

Andrew H. Paterson
Editor

Plant Genetics / Genomics Volume 3

Genetics and Genomics of Cotton



 Springer

Plant Genetics and Genomics: Crops and Models

Volume 3

Plant Genetics and Genomics: Crops and Models

Series Editor: Richard A. Jorgensen

Forthcoming and planned volumes

- Vol. 1 Genomics of Tropical Crop Plants (eds: Paul Moore/Ray Ming)
- Vol. 2 Genetics and Genomics of Soybean (ed: Gary Stacey)
- Vol. 3 Genetics and Genomics of Cotton (ed: Andrew Paterson)
- Vol. 4 Plant Cytogenetics: Genome Structure and Chromosome Function
(eds: Hank Bass/Jim Birchler)
- Vol. 5 Plant Cytogenetics: Methods and Instruction (eds: Hank Bass/Jim
Birchler)
- Vol. 6 Genetics and Genomics of the Rosaceae (eds: Kevin Folta/Sue Gardiner)
- Vol. 7 Genetics and Genomics of the Triticeae (ed: Catherine Feuillet/Gary
Muehlbauer)
- Vol. 8 Genomics of Poplar (ed: Stefan Janssen et al.)

Andrew H. Paterson
Editor

Genetics and Genomics of Cotton

 Springer

Editor

Andrew H. Paterson
University of Georgia
Plant Genome Mapping Laboratory
111 Riverbend Road
Athens, GA 30602
USA
paterson@uga.edu

ISBN 978-0-387-70809-6 e-ISBN 978-0-387-70810-2
DOI 10.1007/978-0-387-70810-2

Library of Congress Control Number: 2008941164

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

springer.com

Preface

A few members of the *Gossypium* (cotton) genus are cultivated for the production of elongated single-celled fibers valued worldwide at about \$20 billion annually at the farm gate, and which sustain one of the world's largest industries (textiles) with an annual worldwide economic impact of about \$500 billion. In a number of ways, cotton production and the textile industry are closely tied to petrochemical usage. Practical application of cotton genomics offers the means to improve both sustainability of crop production, and utilization of bio-based alternatives to petrochemicals, as well as consumer satisfaction with the end product.

The *Gossypium* (cotton) genus also presents unique opportunities to advance our understanding of the natural world. In particular, the evolution of cultivated cottons from their wild ancestors has involved a fascinating series of events that offer scientists the opportunity to dissect the evolution of a novel organ, the 'lint fiber', and also to better understand the roles of polyploidy in generation of biodiversity and in crop productivity and quality.

In this book, advances of the past decade will be summarized and synthesized to elucidate the current state of knowledge of the structure, function, and evolution of the *Gossypium* genome, and progress in the application of this knowledge to cotton improvement. This book will address five broad topics. First, as a backdrop, it is important to understand the naturally-occurring diversity in the genus, its organization and distribution, and its evolutionary history. Of special importance is the formation of a single polyploid from two (among 8) diploid genome types, the radiation of this polyploid, and the independent domestication and improvement of two (among nearly 50) diploid and two (among 5) polyploid species.

Current and future cotton genomics studies are guided by a long history of classical genetics research, as well as nearly two decades of molecular genetics. Natural priorities in cotton improvement, including the genetic control of variation in cotton fiber yield and quality components, disease and pest responses, and abiotic stress responses, will be addressed.

Genomic tools for cotton biology and improvement have expanded dramatically in the past 5 years – a detailed summary of these tools and their early applications is central to this book.

The transition to DNA-based genetics has led to detailed characterization of the cotton genome, solving some mysteries but revealing others in the convoluted course of *Gossypium* evolution and improvement. A few unique features of *Gossypium* biology will be highlighted, illustrating how further study of *Gossypium* promises to advance our understanding of the natural world.

A natural long-term goal in such a diverse and complex genus as *Gossypium* is a synthesis of its evolutionary history and genetic potential into a mechanistic understanding of the key steps that have led to the productivity of cultivated cottons, and the range of diversity among wild relatives. The final chapter will attempt to point the way to such a synthesis, and the central role that genomic approaches will play.

Athens, Georgia, USA

Andrew H. Paterson

Contents

Part I Natural History and Genetic Diversity

Evolution and Natural History of the Cotton Genus	3
Jonathan F. Wendel, Curt Brubaker, Ines Alvarez, Richard Cronn, and James McD. Stewart	
The Worldwide Gene Pool of <i>G. hirsutum</i> and its Improvement.	23
Edward L. Lubbers and Peng W. Chee	
The Worldwide Gene Pool of <i>Gossypium barbadense</i> L. and Its Improvement	53
Richard G. Percy	
The Worldwide Gene Pools of <i>Gossypium arboreum</i> L. and <i>G. herbaceum</i> L., and Their Improvement	69
Venkatesh N. Kulkarni, Basavaraj M. Khadi, Manjula S. Maralappanavar, Lalitadas A. Deshapande, and S. S. Narayanan	

Part II Genomic Tools, Resources and Approaches

<i>Gossypium</i> DNA Markers: Types, Numbers, and Uses	101
Mehboob-ur-Rahman, Yusuf Zafar, and Andrew H. Paterson	
Physical Composition and Organization of the <i>Gossypium</i> Genomes	141
Lifeng Lin and Andrew H. Paterson	
The <i>Gossypium</i> Transcriptome.	157
Joshua A. Udall	
Genetic Engineering of Cotton.	187
Norma L. Trolinder	

Mutagenesis Systems for Genetic Analysis of <i>Gossypium</i>	209
Dick Auld, Ginger G. Light, Mohamed Fokar, Efreem Bechere, and Randy D. Allen	
<i>Gossypium</i> Bioinformatics Resources	227
Alan R. Gingle	
 Part III Bridging Classical and Genomic Investigations of Cotton Biology	
Bridging Classical and Molecular Cytogenetics of <i>Gossypium</i>	257
N'Guessan Olivier Konan, Jean-Pierre Baudoin, Angélique D'Hont, and Guy Mergeai	
Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development	283
Peng W. Chee and B. Todd Campbell	
Bridging Classical and Molecular Genetics of Cotton Disease Resistance	313
Robert J. Wright, Chen Niu, and Bay Nguyen	
Bridging Classical and Molecular Genetics of Abiotic Stress Resistance in Cotton	337
Yehoshua Saranga, Andrew H. Paterson, and Avishag Levi	
Bridging Traditional and Molecular Genetics in Modifying Cottonseed Oil	353
Qing Liu, Surinder Singh, Kent Chapman, and Allan Green	
 Part IV Early Messages	
Genomics of Cotton Fiber Secondary Wall Deposition and Cellulose Biogenesis	385
Candace H. Haigler, Bir Singh, Guirong Wang, and Deshui Zhang	
Responses of the Cotton Genome to Polyploidy	419
Keith L. Adams, Lex Flagel, and Jonathan F. Wendel	
Comparative Genomics of Cotton and <i>Arabidopsis</i>	431
Junkang Rong and Andrew H. Paterson	

Impacts on Agroecosystems of Transgenic Insect and Herbicide Resistance in Cotton	451
Lawrence C. Davis	

Part V Synthesis

Toward Characterizing the Spectrum of Diversity in the <i>Gossypium</i> Genus	483
Andrew H. Paterson	

Index	493
--------------------	-----

Contributors

Keith L. Adams UBC Botanical Garden and Centre for Plant Research and Department of Botany, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada

Randy D. Allen Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK, USA, randy.allen@okstate.edu

Ines Alvarez Real Jardín Botánico de Madrid, CSIC, Plaza de Murillo, 2, E-28014 Madrid, Spain, ines@rjb.csic.es

Dick Auld Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, dick.auld@ttu.edu

Jean-Pierre Baudoin Gembloux Agricultural University (FUSAGx), Unité de Phytotechnie tropicale et d'Horticulture, 2 passage des Déportés, B-5030 Gembloux, Belgium, konan.n@fsagx.ac.be; nguessanolivier@yahoo.fr; baudoin.jp@fsagx.ac.be; mergeai.g@fsagx.ac.be

Efrem Bechere Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, dick.auld@ttu.edu

Curt Brubaker Bayer BioScience N.V., Technologiepark 38, B-9052 Gent, Belgium, curt.brubaker@bayercropsscience.com

B. Todd Campbell Coastal Plains Soil, Water, and Plant Research Center, USDA-ARS, Washington, DC, USA, Todd.Campbell@ars.usda.gov

Kent Chapman Department of Biological Sciences, University of Northern Texas, Denton, TX, USA, chapman@unt.edu

Peng W. Chee Department of Crop and Soil Sciences, University of Georgia, Athens, GA, USA, pwchee@uga.edu

Richard Cronn Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA, rcronn@fs.fed.us

Angélique D'Hont Centre International en Recherche Agronomique pour le Développement (CIRAD), UMR1096, TA 70/03, Avenue Agropolis, 34398 Montpellier Cedex 5, France, angelique.dhont@cirad.fr

Lawrence C. Davis Biochemistry Department, Kansas State University, Manhattan, KS 66506, USA, ldavis@ksu.edu

Lalitadas A. Deshapande Head, Crop Improvement, Central Institute of Cotton Research, Nagpur, Maharashtra, India

Lex Flagel Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

Mohamed Fokar Center for Biotechnology and Genomics, Texas Tech University, Lubbock, TX, USA, Mohamed.fokar@ttu.edu

Alan R. Gingle Plant Genome Mapping Lab, University of Georgia, Athens, GA, USA, agingle@uga.edu

Allan Green CSIRO Plant Industry, Clayton South, VIC, Australia, qing.liu@csiro.au; surinder.singh@csiro.au; allan.green@csiro.au

Candace H. Haigler Department of Crop Science and Department of Plant Biology, North Carolina State University, Raleigh, NC, USA, candace_haigler@ncsu.edu

Basavaraj M. Khadi Central Institute of Cotton Research, Nagpur, Maharashtra, India, cicrngp@rediffmail.com

N'Guessan Olivier Konan Gembloux Agricultural University (FUSAGx), Unité de Phytotechnie tropicale et d'Horticulture, 2 passage des Déportés, B-5030 Gembloux, Belgium, konan.n@fsagx.ac.be; nguessanolivier@yahoo.fr; baudoin.jp@fsagx.ac.be; mergeai.g@fsagx.ac.be

Venkatesh N. Kulkarni Senior Cotton Breeder, JK Agri-Genetics Ltd, 1-10-177, 4th Floor, Varun Towers, Begumpet, Hyderabad – 500 016, India, vnkulkarni@jkseeds.net

Avishag Levi The RH Smith Institute of Plant Science and Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot 76100, Israel

Ginger G. Light Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, ginger.light@ttu.edu

Lifeng Lin Plant Biology Department, University of Georgia, Athens, GA, USA, phill@uga.edu; Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA, paterson@dogwood.botany.uga.edu

Qing Liu CSIRO Plant Industry, Clayton South, VIC, Australia, qing.liu@csiro.au, surinder.singh@csiro.au, allan.green@csiro.au

Edward L. Lubbers Department of Crop and Soil Sciences, University of Georgia, Tifton, Georgia, USA, elubbers@uga.edu

Manjula S. Maralappanavar Agricultural Research Station, Dharwad Farm, Dharwad – 580 007, Karnataka, India, manjula_arsd@yahoo.co.in

Guy Mergeai Gembloux Agricultural University (FUSAGx), Unité de Phytotechnie tropicale et d'Horticulture, 2 passage des Déportés, B-5030 Gembloux, Belgium, konan.n@fsagx.ac.be; nguessanolivier@yahoo.fr; baudoin.jp@fsagx.ac.be; mergeai.g@fsagx.ac.be.

S.S. Narayanan Consultant, JK Agri-Genetics Ltd, 1-10-177, 4th Floor, Varun Towers, Begumpet, Hyderabad – 500 016, India

Bay Nguyen Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, robert.wright@ttu.edu

Chen Niu Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, robert.wright@ttu.edu

Andrew H. Paterson Plant Biology Department, University of Georgia, Athens, GA, USA, phill@uga.edu; University of Georgia, Plant Genome Mapping Laboratory, 111 Riverbend Road, Athens, GA 30602, USA, paterson@dogwood.botany.uga.edu; paterson@dogwood.botany.uga.edu; jkrong@yahoo.com; paterson@uga.edu

Richard G. Percy USDA-ARS, Southern Plains Agricultural Research Center, College Station, Texas, USA, Richard.Percy@ars.usda.gov

Mehboob-ur-Rahman Plant Genomics & Mol Breeding Labs, National Institute for Biotechnology & Genetic Engineering (NIBGE), Faisalabad, Pakistan, mehboob_pbd@yahoo.com

Junkang Rong Plant Genome Mapping Laboratory, The University of Georgia, Athens, GA, USA, jkrong@yahoo.com

Yehoshua Saranga The RH Smith Institute of Plant Science and Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot 76100, Israel, saranga@agri.huji.ac.il

Bir Singh Department of Crop Science, North Carolina State University, Raleigh, NC, USA, bsingh4@ncsu.edu

Surinder Singh CSIRO Plant Industry, Clayton South, VIC, Australia, qing.liu@csiro.au, surinder.singh@csiro.au, allan.green@csiro.au

James McD. Stewart Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR 72701, USA, jstewart@uark.edu

Norma L. Trolinder Texas Tech University & Genes Plus, Lubbock, TX, USA, genes@chipshot.net

Joshua A. Udall Department of Plant and Wildlife Science, Brigham Young University, Provo, Utah, USA, jaudall@byu.edu

Guirong Wang Data Systems Group, Sacramento, CA, USA, guirong2004@yahoo.com

Jonathan F. Wendel Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA, jfw@iastate.edu

Robert J. Wright Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX, USA, robert.wright@ttu.edu; Texas AgriLife Research, Texas A&M University, College Station, TX, USA

Yusuf Zafar National Institute for Genomics & Advanced Biotechnology (NIGAB), NARC, Islamabad, Pakistan, y_zafar@yahoo.com

Deshui Zhang Ventria Bioscience, Fort Collins, CO, USA, dzhang@ventria.com

Evolution and Natural History of the Cotton Genus

Jonathan F. Wendel, Curt Brubaker, Ines Alvarez, Richard Cronn,
and James McD. Stewart

Abstract We present an overview of the evolution and diversity in *Gossypium* (the cotton genus). This framework facilitates insight into fundamental aspects of plant biology, provides the necessary underpinnings for effective utilization of cotton genetic resources, and guides exploration of the genomic basis of morphological diversity in the genus. More than 50 species of *Gossypium* are distributed in arid to semi-arid regions of the tropics and subtropics. Included are four species that independently have been domesticated for their fiber, two each in Africa-Asia and the Americas. *Gossypium* species exhibit extraordinary morphological variation, ranging from trailing herbaceous perennials to ~15 m trees with a diverse array of reproductive and vegetative characteristics. A parallel level of cytogenetic and genomic diversity has arisen during the global radiation of the genus, leading to the evolution of eight groups of diploid ($n = 13$) species (genome groups A through G, and K). Data implicate an origin for *Gossypium* about 5–10 million years ago and a rapid early diversification of the major genome groups. Allopolyploid cottons appear to have arisen within the last 1–2 million years, as a consequence of trans-oceanic dispersal of an A-genome taxon to the New World followed by hybridization with an indigenous D-genome diploid. Subsequent to formation, allopolyploids radiated into three modern lineages, two of which contain the commercially important species *G. hirsutum* and *G. barbadense*.

1 Introduction to *Gossypium* diversity

Because the cotton genus (*Gossypium* L.) is so important to economies around the world, it has long attracted the attention of agricultural scientists, taxonomists, and evolutionary biologists. Accordingly, and notwithstanding the

J.F. Wendel (✉)

Department of Ecology, Evolution, and Organismal Biology, Iowa State University,
Ames, Iowa 50011
e-mail: jfw@iastate.edu

remaining gaps in our knowledge, a great deal is understood about the origin and diversification of the genus. Especially in the last two decades, modern molecular technologies have been brought to bear on such classic questions as the origin of the polyploid species, the relationships among species and species groups, and the origins of the domesticated forms from their wild progenitors.

Perhaps the most striking aspect of this history is that it is so widespread in scope, involving ancient human cultures on several continents and a convergent or parallel plant domestication process from divergent and geographically isolated wild ancestors. This parallel domestication process involved four species, two from the Americas, *G. hirsutum* and *G. barbadense*, and two from Africa-Asia, *G. arboreum* and *G. herbaceum*. In each of these four cases, aboriginal peoples discovered thousands of years ago that the unique properties of cotton fibers made them useful for ropes, textiles and other applications. Each of the domesticated species has its own unique history of domestication, diversification, and utilization, as detailed in many papers describing various stages in the domestication process, the origin of present patterns of genetic diversity, the shape and severity of genetic bottlenecks that accompanied the development of landraces and cultivars, and the influence of recent human history on geographic patterns of cultivation (Brubaker, Bourland and Wendel 1999; Brubaker and Wendel 1994, 2001; Hutchinson 1951, 1954, 1959; Hutchinson, Silow and Stephens 1947; Percy and Wendel 1990; Wendel, Brubaker and Percival 1992; Wendel, Olson and Stewart 1989).

This rich history involved human shaping and molding of naturally occurring diversity that originated through the process of evolutionary diversification over a period of millions of years, a legacy we continue to exploit today through deliberate introgression of alien germplasm from diverse, wild gene pools. Given current threats to global ecosystems, it has never been more critical than it is today to recognize this legacy of speciation and diversification. Toward this end a synopsis of the present understanding of the evolutionary history and taxonomy of *Gossypium* is given, and in the process an entry into the relevant body of literature is provided for the serious reader.

2 Origin of the Genus *Gossypium*

The cotton genus belongs to a small taxonomic tribe, the *Gossypieae*, that includes only eight genera (Fryxell 1968; Fryxell 1979). Four of these genera are small with restricted geographic distributions (Fryxell 1968; Fryxell 1979) including *Lebronnecia* (Marquesas Islands), *Cephalohibiscus* (New Guinea, Solomon Islands), *Gossypoides* (east Africa, Madagascar), and *Kokia* (Hawaii). The tribe also includes four moderately sized genera with broader geographic ranges: *Hampea*, with 21 neotropical species; *Cienfuegosia*, a diverse genus with 25 species from the neotropics and parts of Africa; *Thespesia*, with 17 tropical species; and last but not least, *Gossypium*, the largest and most widely distributed

genus in the tribe with more than 50 species (Fryxell 1992; Stewart, Craven, Brubaker and Wendel 2008; Stewart and Ulloa unpublished).

Molecular phylogenetic analyses have clarified several aspects of the evolutionary history of the tribe (Cronn, Small, Haselkorn and Wendel 2002; Seelanan, Schnabel and Wendel 1997). Most important has been the demonstration that the group of species recognized as belonging to *Gossypium* do in fact constitute a single natural lineage, despite their global distribution and extraordinary morphological and cytogenetic diversity. A second discovery has been the identity of the closest relatives of *Gossypium*, i.e., the African-Madagascan genus *Gossypioides* and the Hawaiian endemic genus *Kokia*. These latter genera may thus be used as phylogenetic outgroups for studying evolutionary patterns and processes within *Gossypium*. A third insight concerns the temporal component to the genealogy, which is evident in sequence divergence data that serve as a proxy for time. Using this “molecular clock”, Seelanan, Schnabel and Wendel (1997) suggested that *Gossypium* branched off from *Kokia* and *Gossypioides* approximately 12.5 million years ago (mya), in agreement with a later, more extensive data set based on 10 different nuclear genes (Cronn, Small, Haselkorn and Wendel 2002). Thus, *Gossypium* appears to have diverged from its closest relatives during the Miocene, perhaps 10–15 mya, subsequently spreading around the world via trans-oceanic dispersal to acquire its modern geographic range.

3 Diversification and Speciation of the Diploid Cotton Species

From its origin millions of years ago, *Gossypium* underwent speciation and diversification, achieving a nearly worldwide distribution with several primary centers of diversity in the arid or seasonally arid tropics and subtropics (Table 1). Species-rich regions include Australia, especially the Kimberley region in NW Australia, the Horn of Africa and southern Arabian Peninsula, and the western part of central and southern Mexico. Recognition of these groups of related species and their individual constituents reflects accumulated scientific understanding that emerged from basic plant exploration and taxonomic and evolutionary study. The taxonomy of the genus has been well-studied (Cronn, Small, Haselkorn and Wendel 2002; Fryxell 1979, 1992; Hutchinson, Silow and Stephens 1947; Saunders 1961; Seelanan, Schnabel and Wendel 1997; Watt 1907), with the most modern and widely followed taxonomic treatments being those of Fryxell (1979; 1992) in which species are grouped into four subgenera and eight sections (Table 1). This classification system is based primarily on morphological and geographical evidence, although most infra-generic alignments are congruent with cytogenetic and molecular data sets as well (see Chapter 11 by Mergeai, this volume).

At present, *Gossypium* includes approximately 50 species (Fryxell 1992), but remarkably, new species continue to be discovered (Fryxell 1992; Stewart, Craven, Brubaker and Wendel 2008; Ulloa, Stewart, Garcia, Godoy, Gaytán

Table 1 Diversity and geographic distribution of the major lineages of *Gossypium*. Genomic placements of species enclosed by parentheses are yet to be determined

Genome group	Number of species	Recognized species	Geographic distribution
A	2	<i>G. arboreum</i> , <i>G. herbaceum</i>	Africa, Asia
B	3 (4)	<i>G. anomalum</i> <i>G. triphyllum</i> <i>G. capitiviridis</i> (<i>G. trifurcatum</i>)	Africa, Cape Verde Islands
C	2	<i>G. sturtianum</i> <i>G. robinsonii</i>	Australia
D	13(14)	<i>G. thurberi</i> <i>G. armourianum</i> <i>G. harknessii</i> <i>G. davidsonii</i> <i>G. klotschianum</i> <i>G. aridum</i> <i>G. raimondii</i> <i>G. gossypoides</i> <i>G. lobatum</i> <i>G. trilobatum</i> <i>G. laxum</i> <i>G. turneri</i> <i>G. schwendimanii</i> (<i>G. sp.nov.</i>)	Primarily Mexico; also Peru, Galapagos Islands, Arizona
E	5 (9)	<i>G. stocksii</i> <i>G. somalense</i> <i>G. areysianum</i> <i>G. incanum</i> <i>G. trifurcatum</i> (<i>G. benidirensense</i>) (<i>G. bricchettii</i>) (<i>G. vollesenii</i>) (<i>G. trifurcatum</i>)	Arabian Peninsula, Northeast Africa, Southwest Asia
F	1	<i>G. longicalyx</i>	East Africa
G	3	<i>G. bickii</i> <i>G. australe</i> <i>G. nelsonii</i>	Australia
K	12	<i>G. anapoides</i> <i>G. costulatum</i> <i>G. cunninghamii</i> <i>G. enthyale</i> <i>G. exiguum</i> <i>G. londonderriense</i> <i>G. marchantii</i> <i>G. nobile</i> <i>G. pilosum</i> <i>G. populifolium</i> <i>G. pulchellum</i> <i>G. rotundifolium</i>	NW Australia, Cobourg Peninsula, NT

Table 1 (continued)

Genome group	Number of species	Recognized species	Geographic distribution
AD	5	<i>G. hirsutum</i> <i>G. barbadense</i> <i>G. tomentosum</i> <i>G. mustelinum</i> <i>G. darwinii</i>	New World tropics and subtropics, including Hawaii

and Acosta 2006). The genus is extraordinarily diverse; species morphologies range from fire-adapted, herbaceous perennials in NW Australia to trees in SW Mexico that escape the dry season by dropping their leaves. Corolla colors span a rainbow of blue to purple (*G. triphyllum*), mauves and pinks (“Sturt’s Desert Rose”, *G. sturtianum*, is the official floral emblem of the Northern Territory, Australia), whites and pale yellows (NW Australia, Mexico, Africa-Arabia) and even a deep sulphur-yellow (*G. tomentosum* from Hawaii). Seed coverings range from nearly glabrous to the naked eye (e.g., *G. klotzschianum* and *G. davidsonii*), to short stiff, dense, brown hairs that aid in wind-dispersal (*G. australe*, *G. nelsonii*), to the long, fine white fibers that characterize highly improved forms of the four cultivated species. There are even seeds that produce fat bodies to facilitate ant-dispersal (Seelanan, Brubaker, Stewart, Craven and Wendel 1999).

As the genus diversified and spread, it underwent extensive chromosomal evolution (see Chapter 11 by Konan et al. and Chapter 11 by Mergeai, this volume). Although all diploid species share the same chromosome number ($n = 13$), there is more than three-fold variation in DNA content per genome (Hendrix and Stewart 2005). Chromosome morphology is similar among closely related species, and this is reflected in the ability of related species to form hybrids that display normal meiotic pairing and sometimes high F_1 fertility. In contrast, crosses among more distant relatives may be very difficult to effect, and those that are successful are characterized by meiotic abnormalities. The collective observations of pairing behavior, chromosome sizes, and relative fertility in interspecific hybrids led to the designation of single-letter genome symbols (Beasley 1941) for related clusters of species. Presently, eight diploid genome groups (A through G, plus K) are recognized (Endrizzi, Turcotte and Kohel 1985; Stewart 1995). This cytogenetic partition of the genus is largely congruent with taxonomic and phylogenetic divisions. A brief introduction to the major groups of diploid cotton species follows.

3.1 Australian Species

Australian cottons (subgenus *Sturtia*) comprise 16 named species as well as a new species whose description is in press (Stewart, Craven, Brubaker and

Wendel 2008). Collectively, these taxa comprise the C-, G-, and K-genome groups, with two, three, and twelve species, respectively. These three groups of species are implicated by DNA sequence data (Liu, Brubaker, Green, Marshall, Sharp and Singh 2001; Seelanan, Brubaker, Stewart, Craven and Wendel 1999; Seelanan, Schnabel and Wendel 1997) to be natural lineages, consistent with their formal alignments into the taxonomic sections *Sturtia* (C-genome), *Hibiscoidea* (G-genome), and *Grandicalyx* (K-genome). Relationships among the three groups, however, remain unclear. Some data place *G. robinsonii* as basal within the entire assemblage of Australian species (DeJooide and Wendel 1992), suggesting that radiation of *Gossypium* in Australia proceeded eastward from the westernmost portion of the continent. Whether this basal position will withstand the scrutiny of other data sets is an open question, as the most recent analyses (Liu, Brubaker, Mergeai, Cronn and Wendel 2001; Seelanan, Brubaker, Stewart, Craven and Wendel 1999; Seelanan, Schnabel and Wendel 1997) are equivocal in this regard.

With respect to the taxonomy within each of the three Australian genome groups, there is little uncertainty for the C- and G-genome groups, as these are well represented in collections and have been thoroughly studied (DeJooide and Wendel 1992; Fryxell 1979, 1992; Liu, Brubaker, Mergeai, Cronn and Wendel 2001; Seelanan, Brubaker, Stewart, Craven and Wendel 1999; Seelanan, Schnabel and Wendel 1997; Wendel, Stewart and Rettig 1991). Much less certain is the taxonomy of the K-genome species, which are all placed in section *Grandicalyx*. Collecting expeditions to the Kimberley area have enhanced our understanding of diversity within the group and have resulted in the discovery of at least seven new species (Fryxell, Craven and Stewart 1992; Stewart, Craven, Brubaker and Wendel, 2008). These unusual species have a distinctive geography, morphology and ecology, and exhibit a syndrome of features that are characteristic of fire-adaptation. In particular, they are herbaceous perennials with a bi-seasonal growth pattern whereby vegetative growth dies back during the dry season to underground rootstocks that initiate a new cycle of growth with the onset of the next wet season or following a fire. Species in section *Grandicalyx* have flowers that are upright when open but that become pendant following pollination so that at maturity, the capsules release sparsely haired, ant-dispersed seeds that bear elaiosomes (fat bodies) to aid in attracting ants. Many of these species are poorly represented in collections and not well understood taxonomically. Molecular phylogenetic analyses have yielded conflicting results regarding interspecific relationships in this group (Liu, Brubaker, Mergeai, Cronn and Wendel 2001; Seelanan, Brubaker, Stewart, Craven and Wendel 1999).

3.2 African-Asian Species

Fourteen species from Africa and Arabia are recognized in the most recent taxonomic treatment of the genus (Fryxell 1992), collectively in subgenus

Gossypium. The taxonomic section *Gossypium* contains four subsections, whereas section *Serrata* contains only *G. trifurcatum* from the desert area of eastern Somalia. The presence of dentate leaves raised the possibility that it may not belong in *Gossypium*, but recent molecular work clearly established this poorly known entity as a *bona fide* if unusual cotton species (Rapp, Alvarez and Wendel 2005). This latter example underscores the provisional nature of much of the taxonomy of the African-Arabian species of *Gossypium*, which are sorely in need of basic plant exploration and systematic study. Within section *Pseudopambak*, species recognition and definition are in some cases based on limited herbarium material (e.g., *G. benadirensis*, *G. bricchettii*, *G. vollesenii*) and seeds have not been collected. Consequently, no analyses have been conducted on cytogenetic characteristics nor molecular phylogenetic affinities.

From a cytogenetic standpoint, the African-Arabian species exhibit considerable diversity, collectively accounting for four of the eight genome groups (A-, B-, E-, and F-). The A-genome group comprises the two cultivated cottons of subsection *Gossypium*, *G. arboreum* and *G. herbaceum*. Three African species in subsection *Anomala* (*G. anomalum*, *G. captis-viridis* and *G. triphyllum*) comprise the B genome. *Gossypium trifurcatum* may also belong to the B genome, but this has not been established. The sole F-genome species, *G. longicalyx*, is cytogenetically distinct (Phillips 1966), morphologically isolated (Fryxell 1971; Fryxell 1992) and, according to Fryxell (1979), is perhaps adapted to more mesic conditions than any other diploid *Gossypium* species. The remaining African-Asian species, those of subsection *Pseudopambak*, are considered to possess E-genomes, although this has yet to be verified.

3.3 American Diploid Species

Subgenus *Houzingenia* contains two sections and six subsections, whose species collectively represent the New World D-genome diploids. These species have been more thoroughly studied than most, and consequently their taxonomy is reasonably well-understood. This subgenus has also received considerable phylogenetic attention (Álvarez, Cronn and Wendel 2005; Cronn, Small, Haselkorn and Wendel 2002; Cronn, Small, Haselkorn and Wendel 2003; Small and Wendel 2005; Wendel and Albert 1992), which provides support for the naturalness of most of the subsections. Evolutionary relationships among the apparently natural subsections are less certain, however (Álvarez, Cronn and Wendel 2005), although available evidence suggests that *G. gossypioides* is basal-most within the subgenus (Cronn, Small, Haselkorn and Wendel 2003).

Twelve of the 14 D-genome diploid species are endemic to western Mexico, thus this area is the center of diversity of the D genome. Likely, the lineage became established and initially diversified in this region. Later range extensions are inferred to have arisen from relatively recent (probably Pleistocene) long-distance dispersals, leading to the evolution of endemics in Peru (*G. raimondii*) and the Galapagos Islands (*G. klotzschianum*).

4 Origin and Diversification of the Polyploid Cottons

Classic cytogenetic investigations demonstrated that the American tetraploid species are allopolyploids containing two resident genomes, an A-genome from Africa or Asia, and a D-genome similar to those found in the American diploids (Beasley 1940; Denham 1924; Harland 1940; Skovsted 1934, 1937; Webber, 1935). Additional support for the hypothesis of an allopolyploid origin of the American tetraploids emerged in subsequent decades from numerous sources of evidence, including the synthesis of additional experimental allotetraploids (Stewart 1995). This history and evidence is detailed in Endrizzi, Turcotte and Kohel (1985), Wendel and Cronn (2003) and Chapter 11 by Mergeai, this volume.

When did allopolyploid cottons first form, and how did this happen given that the two diploid genomes involved (A and D) presently exist in species from different hemispheres? This question was a classic botanical mystery for over half a century (see Endrizzi, Turcotte and Kohel 1985), a mystery at least partially solved through the recent use of molecular technologies. With respect to the first part of the question, that of “when”, gene sequence data convincingly demonstrate that allopolyploid *Gossypium* originated prior to the evolution of modern humans but relatively recently in geological terms, perhaps 1–2 mya, or in the mid-Pleistocene (Cronn, Small, Haselkorn and Wendel 2002; Seelanan, Schnabel and Wendel 1997; Senchina, Alvarez, Cronn, Liu, Rong, Noyes, Paterson, Wing, Wilkins and Wendel 2003). With respect to the second part of the question, that of polyploid parentage, it is now clear that both extant A-genome species (*G. arboreum*, *G. herbaceum*) are equally divergent from the A-genome of allopolyploid cottons and that the closest living relative of the progenitor D-genome donor is *G. raimondii* (Endrizzi, Turcotte and Kohel 1985; Wendel and Cronn 2003). One aspect of the history of the polyploid cottons that has become clear is that they all contain an A-genome cytoplasm, and most likely from a single source (Galau and Wilkins 1989; Small and Wendel 1999; Wendel 1989). Studies using nuclear (bi-parentally inherited) genes lead to the same conclusion. Hence, evidence indicates that natural allopolyploid cottons all derive from a single lineage.

Given a Pleistocene origin for allopolyploid cotton species, one may infer that their morphological diversification and spread must have been relatively rapid following polyploidization. At present, five allopolyploid species are recognized. *Gossypium darwinii* is native to the Galapagos Islands, where it may form large and continuous populations in some areas (Percy and Wendel 1990). *Gossypium tomentosum*, from the Hawaiian Islands, has a more diffuse population structure, occurring mostly as scattered individuals and small populations on several islands (DeJooe and Wendel 1992). A third allopolyploid, *G. mustelinum*, is an uncommon species restricted to a relatively small region of northeast Brazil (Wendel, Rowley and Stewart 1994). In addition to these three truly wild species, there are two cultivated species (*G. barbadense* and

G. hirsutum), each of which has a large indigenous range, collectively encompassing a wealth of morphological forms that span the wild-to-domesticated continuum (Brubaker and Wendel 1993, 1994, 2001; Fryxell 1979; Hutchinson 1951; Percy and Wendel 1990). *Gossypium hirsutum* is widely distributed in Central and northern South America, the Caribbean, and even reaches distant islands in the Pacific (Solomon Islands, Marquesas). *Gossypium hirsutum* is thought to have a more northerly distribution than *G. barbadense*, with wild populations occurring as far north (27°38'N) as Tampa Bay (Stewart, personal observation). However, recently Stewart and Bertoni (2007 unpublished) collected a wild population of *G. hirsutum* in the Chaco state of Presidente Hayes, Paraguay (~22°S). *Gossypium barbadense* has a more southerly indigenous range, centered in the northern third of South America but with a large region of range overlap with *G. hirsutum* in the Caribbean.

Consideration of the distribution of the allopolyploid species suggests that polyploidy led to the invasion of a new ecological niche. Fryxell (1965, 1979) noted that in contrast to the majority of diploid species, allopolyploid species typically occur in coastal habitats, at least those forms that arguably are truly wild. Two species, both island endemics (*G. darwinii* and *G. tomentosum*), are restricted to near coastlines, and for two others (*G. barbadense* and *G. hirsutum*), wild forms occur in littoral habitats ringing the Gulf of Mexico, northwest South America, and even on distant Pacific Islands. Fryxell (1965, 1979) speculated that following initial formation, adaptation of the newly evolved allopolyploid to littoral habitats enabled it to exploit the fluctuating sea levels that characterized the Pleistocene. This ecological innovation is envisioned to have facilitated initial establishment of the new polyploid lineage, and also may have provided a means for the rapid dispersal of the salt-water tolerant seeds. The recent discovery of *G. hirsutum* in Paraguay does not negate this hypothesis, in that the population was located next to an area of seasonal water accumulation suspected of being saline due to the many years of water accumulation followed by evaporation. The timing and source of the initial introduction of cotton to this area are unknown, but could involve bird dispersal.

5 Phylogenetic Relationships in the Genus

A genealogical framework for the genus has been provided by multiple molecular phylogenetic investigations (reviewed in Wendel and Cronn 2003). Each of these studies shows that genealogical lineages of species are congruent with genome designations and geographical distributions. Cytogenetic studies (reviewed in Chapter 11, this volume) further support this conclusion. Accordingly, each genome group corresponds to a single natural lineage, and in most cases, these lineages are also geographically cohesive. This information is summarized in a depiction of our present understanding of relationships (Fig. 1).

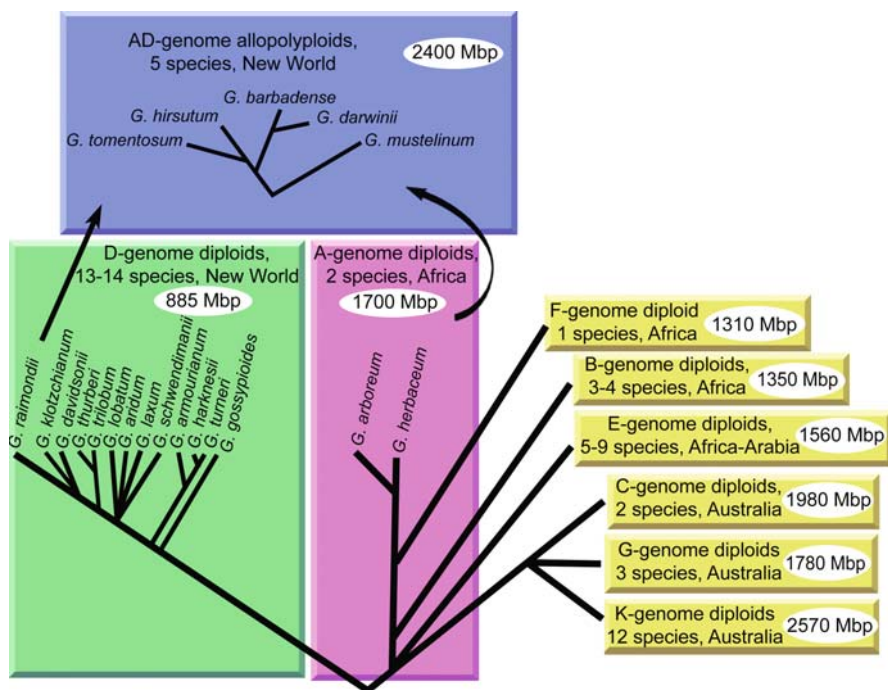


Fig. 1 Evolutionary history of *Gossypium*, as inferred from multiple molecular phylogenetic data sets. The closest relative of *Gossypium* is a lineage containing the African-Madagascan genus *Gossypioides* and the Hawaiian endemic genus *Kokia*. Following its likely origin 5–10 mya, *Gossypium* split into three major diploid lineages: the New World clade (D-genome); the African-Asian clade (A-, B-, E- and F-genomes); and the Australian clade (C-, G-, and K-genomes). This global radiation involved several trans-oceanic dispersal events and was accompanied by morphological, ecological, and chromosomal differentiation (2C genome sizes shown in white ellipses). Interspecific hybridization is implicated in the evolution of approximately one-fourth of the genus. Allopolyploid cottons formed following trans-oceanic dispersal of an A-genome diploid to the Americas, where the new immigrant underwent hybridization, as female, with a native D-genome diploid similar to modern *G. raimondii*. Polyploid cotton probably originated during the Pleistocene (1–2 mya), with the five modern species representing the descendants of an early and rapid colonization of the New World tropics and subtropics (See Color Insert)

Several aspects of the phylogenetic history of *Gossypium* bear highlighting. *First*, there exist four major lineages of diploid species corresponding to three continents: Australia (C-, G-, K-genomes), the Americas (D-genome), and Africa/Arabia (two lineages: one comprising the A-, B-, and F-genomes, and a second containing the E-genome species). *Second*, the earliest divergence event in the genus separated the New World D-genome lineage from the ancestor of all Old World taxa. Thus, New World and Old World diploids are phylogenetic sister groups. Following this basal-most split in the genus, cottons comprising the Old World lineage divided into three groups, namely, the Australian cottons

(C-, G-, and K-genome species), the African-Arabian E-genome species, and the African A-, B-, and F-genome cottons. *Third*, the African F-genome clade, which consists of the sole species *G. longicalyx*, is definitively diagnosed as sister to the A-genome species. This identifies the wild forms most closely related to those first domesticated in the A-genome species *G. arboreum* and *G. herbaceum*. Because this relationship is revealed, prospects are raised for ultimately understanding the genetic basis of the origin of useful, long lint. *Fourth*, the major lineages of *Gossypium* were established in relatively rapid succession shortly after the genus originated and diverged from the *Kokia-Gossypoides* clade. The evolutionary picture thus envisioned is that there was a rapid and global radiation early in the history of the genus, with temporally closely spaced divergence events.

Allopolyploid cottons, including the two species of great commercial importance today (*G. hirsutum* and *G. barbadense*) are implicated to have formed following a biological reunion of two genomes (A and D) that descended from the earliest split in the genus. That is, the two constituent genomes of allopolyploid cotton evolved first in different hemispheres and diverged for millions of years, in isolation from one another. Allopolyploid cottons thus contain duplicated but slightly divergent copies of most genes. Senchina et al. (2003) studied 48 pairs of these duplicates, and showed that on average, there is about 3–4% sequence divergence between copies, although there is considerable variation about this mean.

Consideration of the phylogeny of Fig. 1 in a temporal context and in light of plate tectonic history leads to an inference of multiple intercontinental dispersal events and other episodes of trans-oceanic travel during the evolutionary history of the *Gossypium*. These include at least one dispersal between Australia and Africa, another to the Americas (probably Mexico) leading to the evolution of the D-genome diploids, and a second, much later colonization of the New World by the A-genome ancestor of the AD-genome allopolyploids. Long-distance dispersal played a role not only in diversification of major evolutionary lines but also in speciation within *Gossypium* genome groups. Examples include dispersals from southern Mexico to Peru (*G. raimondii*), from northern Mexico to the Galapagos Islands (*G. klotzschianum*), from western South America to the Galapagos Islands (*G. darwinii*), from Africa to the Cape Verde Islands (*G. capitis-viridis*), and from the neotropics to the Hawaiian Islands (*G. tomentosum*).

6 Seed and Fiber Diversity and Seed Dispersal

An additional perspective offered by the phylogeny of the genus (Fig. 1) concerns the evolutionary history of seeds and their associated single-celled epidermal trichomes, known as hairs in the wild species and lint and fuzz fibers in the cultivated taxa. The seeds and their coverings are extraordinarily

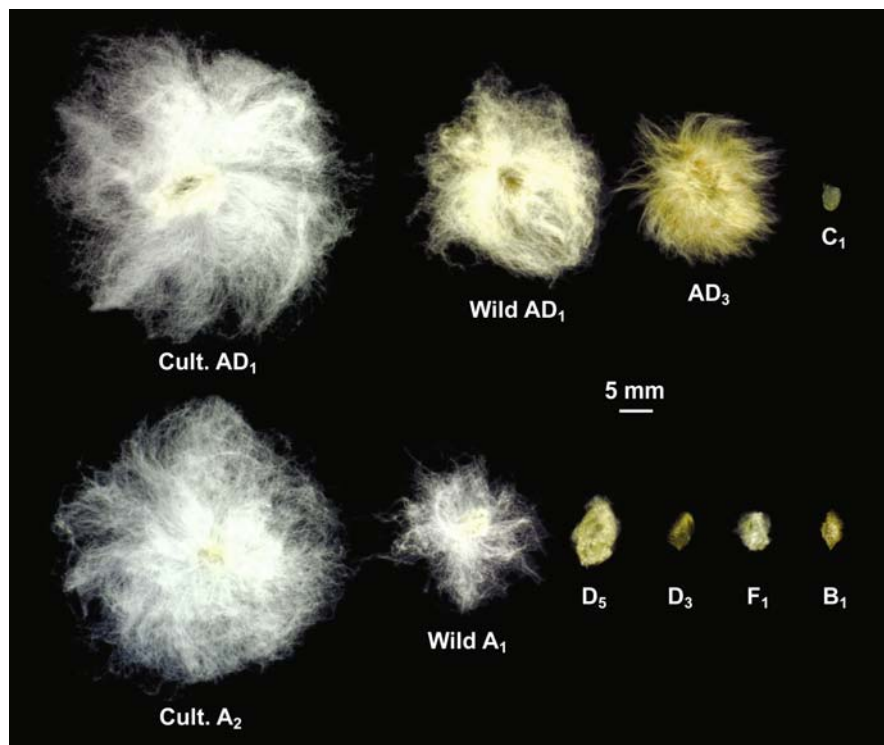


Fig. 2 Representative seed and trichome diversity in *Gossypium*. Seed and trichome size and morphology are exceedingly variable in the genus. Most wild species have relatively small seeds (<5 mm in any dimension) with equally short fibers. Long (spinnable) fiber evolved only once, in the ancestor of modern A-genome cottons, which subsequently donated this capacity to modern tetraploid species, including the commercially important *G. hirsutum* and *G. barbadense*, at the time of allopolyploid formation in the mid-Pleistocene. See text for additional detail. Key to species: Cult. AD₁ = *G. hirsutum* TM1; Wild AD₁ = *G. hirsutum* Tx2094 from the Yucatan Peninsula; AD₃ = *G. tomentosum* WT936 from Hawaii; C₁ = *G. sturtianum* C₁-4 from Australia; Cult. A₂ = *G. arboreum* AKA8401; Wild A₁ = *G. herbaceum* subsp. *africanum* from Botswana; D₅ = *G. raimondii* from Peru; D₃ = *G. davidsonii* D_{3d}-32 from Baja California; F₁ = *G. longicalyx* F₁-3 from Tanzania; B₁ = *G. anomalum* B₁-1 from Africa (See Color Insert)

diverse in *Gossypium*, as shown in Fig. 2, which shows a sampling of the different lineages of wild species and two cultivated species (*G. arboreum* and *G. hirsutum*) for comparison. Both wild and cultivated cottons produce fiber on the seed coat, but there are striking morphological and structural differences between these fibers, the most obvious of which is their size. Some D-genome species (*G. thurberi*, *G. trilobum*, *G. davidsonii*, and *G. klotzschianum*) do not possess obvious seed hairs, but they actually are present as developmentally repressed structures not visible to the unaided eye (Applequist, Cronn and Wendel 2001). Similarly, the three D-genome species of subsection *Cauducibracteolata* give the appearance of being hairless but, in fact,

have seed hairs that are tightly appressed to the seed. Cultivated lint fiber is a single cell of almost pure cellulose, which elongates during development, depending on the species, to a final length of two to six centimeters (Kim and Triplett 2001). The wild-type fiber cell is composed mostly of cellulose and suberin, and elongates to less than one centimeter (Ryser and Holloway 1985; Applequist, Cronn and Wendel 2001). These and many other differences (Fryxell 1979; Hutchinson and Stephens 1945) reflect both natural evolutionary processes as well as human-mediated selection during domestication. The duration of the elongation phase and the timing of onset of secondary wall synthesis appear to be key determinants of the final length of the fiber in both wild and cultivated plants (Applequist, Cronn, and Wendel, 2001). Phylogenetic analysis of growth rates has shown that the evolutionary innovation of prolonged elongation arose in the F-genome/A-genome lineage, which may have facilitated the original domestication of the A-genome cottons. This trait of prolonged elongation was passed on to the allopolyploids, which in turn was a key component of their eventual domestication (Applequist, Cronn, and Wendel 2001). Relatively little is known about the developmental and genetic underpinnings of the diverse morphologies illustrated in Fig. 2, but an improved understanding of the changes that occurred during evolution and domestication may have implications for crop improvement.

In nature, seed dispersal in *Gossypium* often follows a “shaker” model, where erect, mature capsules dehisce along the sutures and the seeds are distributed near the parental plant as wind shakes the branches. This likely is the ancestral method of seed dispersal in the genus, as it occurs in all species of the B-, C-, E-, F-, and D- (in part) genome groups and in one species (*G. bickii*) of the G-genome group. The A-, AD-, some D-genome species, and G- and K-genome species have evolved other mechanisms of seed dispersal, with perhaps the most noteworthy innovation being the fat bodies to facilitate ant dispersal on the seeds of the K-genome species (Seelanan, Brubaker, Stewart, Craven and Wendel 1999). The development of spinnable fibers apparently has occurred only once in the history of *Gossypium*, in the ancestor of the two A-genome species which became the progenitor of the A-subgenome of the tetraploids.

The evolutionary “purpose” of epidermal seed hairs, or trichomes, is a matter of speculation. Fryxell (1979) suggested that elongated fibers aid dispersal by birds, an hypothesis that gains credibility by observations by one of us (J. Stewart) of a bird’s nest in NW Puerto Rico that contained numerous seeds of feral *G. hirsutum*, as well as a collection of *G. darwinii* from a finch’s nest in the Galapagos Islands (JFW, unpubl.). One might also speculate that fibers serve to inhibit germination unless there is sufficient moisture to saturate the fibers; should germination occur following a light rain, there might not be sufficient water for subsequent survival of the seedling. In this respect the waxy coating of the fibers would repel water, to a point, and thus prevent premature germination. A related possibility is that seed hairs function as “biological incubators” to facilitate germination only when ecological conditions are

appropriate, by recruiting particular microbial communities under appropriate moisture regimes. Finally, two G-genome species (*G. australe*, *G. nelsonii*) have evolved stiff straight seed hairs that facilitate wind dispersal in that the stiffening hairs function to extrude seeds from the locules of the dehiscent capsule.

7 Genetic Diversity in the Cultivated Cottons

Domestication has altered the genetic composition of the four cultivated species in comparison to their wild relatives. As is typical of most wild *Gossypium* species, the pre-domesticated cotton species probably had restricted distributions (Brubaker, Bourland and Wendel 1999), although we cannot be certain of this because post-domestication, human-mediated germplasm diffusion has partially (*G. barbadense* and *G. hirsutum*) or completely (*G. herbaceum* and *G. arboreum*) obscured the geographic and genetic origins of each species. It is clear, however, that the indigenous ranges of the four cotton species expanded considerably as cultigens. Eventually this brought *G. barbadense* and *G. hirsutum* into sympatry in the New World, and in the Old World the ranges of *G. herbaceum* and *G. arboreum* also intersected. The consequence was inter-specific gene flow prior to the initiation of deliberate bi-specific breeding programs (e.g., the Acala and Pima cultivar families). Among indigenous New World cotton populations, natural introgression patterns appear to be biased toward the transfer of *G. barbadense* alleles into *G. hirsutum*, particularly in the Caribbean where the ranges of the two species overlap to the greatest extent (Wendel, Brubaker and Percival 1992; Brubaker and Wendel 1994; Brubaker, Koontz and Wendel 1993; Percy and Wendel 1990). Among the modern elite cultivars, however, the converse is more prevalent. Estimates based on RFLP data suggest that up to 8.9% of alleles in modern *G. barbadense* cultivars may have originated from *G. hirsutum* (Wang, Dong and Paterson 1995). The Old World cotton species co-occur in India, northern Africa and China, and there is evidence for low levels of interspecific gene flow, which may explain in part the considerable morphological overlap between the two species (Wendel, Olsen and Stewart 1989).

While pre- and post-domestication interspecific gene flow may have mingled the genetic composition of cotton species to some extent, the genetic winnowing that accompanied human selection progressively narrowed the total available genetic diversity. In the earliest stages of domestication, appreciable levels of genetic diversity were captured among geographic foci of agronomic development, often recognized as “races” (Hutchinson 1951; Hutchinson 1954; Hutchinson and Ghose, 1937; Hutchinson 1962; reviewed by Brubaker, Bourland and Wendel 1999), but the modern elite cultivars now contain only a fraction of this original diversity. As facultative self-pollinators, cotton species tolerate inbreeding, and the most recent stages of domestication accelerated the progressive purge of alleles, resulting in modern elite cultivars with reduced levels of genetic diversity.

Table 2 Comparative genetic diversity estimates for cotton species

Description	Genome	No of Loci	Percent		Mean # Alleles per Locus (<i>A</i>)	Mean Panmictic Heterozygosity (<i>H_T</i>)	Reference
			Polymorphic Loci (<i>P</i>)	Loci (<i>P</i>)			
449 Plant Species			50		1.96	0.15	Hamrick & Godt 1990
Endemics			40		1.80	0.10	Hamrick & Godt 1990
Island endemics			25		1.32	0.06	Hamrick & Godt 1990
<i>G. arboreum</i>	feral, commensal, modern improved cultivars	40	28		2.36	0.066	Wendel <i>et al.</i> , 1989
<i>G. herbaceum</i>	wild, feral, commensal, modern improved cultivars	40	25		2.00	0.039	Wendel <i>et al.</i> , 1989
<i>G. barbadense</i>	wild, feral, commensal, modern improved cultivars	59	41		1.69	0.06	Percy & Wendel 1990
<i>G. darwinii</i>	modern improved cultivars	59	36		1.39	0.058	Wendel & Percy 1990
	wild island endemic	59	27		1.34	0.062	
<i>G. hirsutum</i>	wild, feral, commensal, modern improved cultivars	50	60		2.30	0.163	Wendel <i>et al.</i> , 1992
<i>G. mustelinum</i>	modern improved cultivars	50	28		1.28	0.056	Wendel <i>et al.</i> , 1994
	rare Brazilian endemic	50	12		1.14	0.08	
<i>G. tomentosum</i>	wild island endemic	50	22		1.24	0.033	DeJooode & Wendel 1992

Despite the ready availability of molecular markers over the past ten years, the most informative set of comparative data regarding genetic diversity of *Gossypium* species can be found in a series of early isozyme studies (Table 2). The numbers indicate that the wild tetraploid species have levels of genetic diversity consistent with their occurrence as island endemics (*G. darwinii* and *G. tomentosum*) or their rarity (*G. mustelinum*). To the extent that the progenitor populations of the cotton species were similarly restricted in distribution, they might have been similarly depauperate genetically. The cotton species as they currently exist, however, contain appreciable albeit unremarkable levels of genetic diversity relative to plants in general (Table 2). Of the four cotton species, *G. hirsutum* contains the highest levels of genetic diversity, and it is here that the genetic consequences of domestication are most evident, as the modern upland cultivars contain only half of the genetic diversity available. *Gossypium barbadense* cultivars, having undergone less intensive selection, retain a greater proportion of the diversity present in the species as a whole. More recent molecular marker studies (using RFLPs, RAPDs, AFLPs, or SSRs), while not directly comparable and while focused on *G. hirsutum*, have validated these initial genetic “portraits” (Abdalla, Reddy, El-Zik and Pepper 2001; Brubaker and Wendel 1994; Hawkins, Pleasants and Wendel 2005; Iqbal Reddy, El-Zik and Pepper 2001; Khan, Hussain, Askari, Stewart, Malik and Zafar 2000; Lacape, Dessauw, Rajab, Noyer and Hau 2007; Liu, Cantrell, McCarty and Stewart 2000; Liu, Guo, Lin, Nie and Zhang 2005; Pillay and Myers 1999; Rungis, Llewellyn, Dennis and Lyon 2005), with paucity of genetic diversity among the modern Upland cultivars as a recurring theme (Lacape, Dessauw, Rajab, Noyer, and Hau 2007). One practical consequence of this, exacerbated by the presence of a genome that contains 26 chromosomes, is that marker-assisted selection in Upland cotton breeding is less advanced than it is for other crops. Accordingly, geneticists continue to exploit the more accessible interspecific genetic differences between *G. barbadense* and *G. hirsutum* and between *G. arboreum* and *G. herbaceum* (Ulloa, Brubaker and Chee 2007).

8 Diversity and Crop Improvement

One of the opportunities that arises from an appreciation of the evolutionary history and diversity of *Gossypium* is that of deliberate introgression of useful genes from various wild sources of germplasm. It is a truism that the morphological and ecological breadth encompassed by the wild species must have parallels in the underlying genes that control, for example, physiology, chemistry, disease resistance traits, and fiber characteristics. The wild species of cotton, consequently, represent an ample genetic repository for potential exploitation by cotton breeders. Although these wild species remain a largely untapped genetic resource, there are many examples of their inclusion in breeding programs (reviewed by McCarty and Percy 2001). Cotton improvement

programs have exploited diploid species for genes for fiber strength, disease resistance, cytoplasmic male sterility and fertility restoration, whereas genes for disease resistance, nectariless, and glandless cotton have been deliberately introduced from wild and feral tetraploids. These genetic enhancements, involving intentional interspecific introgression from a minimum of two allopolyploid and four diploid *Gossypium* species (Meredith 1991; Meyer 1974; Narayanan, Singh and Varma 1984; Niles and Feaster 1984), were obtained through classical genetic and plant breeding approaches. Further exploitation of wild *Gossypium* and more phylogenetically distant sources of germplasm will employ these traditional methods as well as genetic engineering (Stewart 1995). This promise is being realized, as efficient transformation systems have been developed (see Chapter 8 by Trolinder, this volume). Herbicide tolerant, insect resistant, and fiber-modified transgenic cultivars are now widespread, accounting for the majority of cotton grown in the world today (see Chapter 19 by Davis, this volume).

Acknowledgments We are grateful to C. Grover and G. Courtney for help with the figures and to the funding agencies that supported much of the work synthesized here. This includes the National Science Foundation and the USDA in the US, and CSIRO in Australia.

References

- Abdalla, A.M., Reddy, O.U.K., El-Zik, K.M. and Pepper, A.E. (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor. Appl. Genet.* 102, 222–229.
- Álvarez, I., Cronn, R. and Wendel, J.F. (2005) Phylogeny of the New World diploid cottons (*Gossypium* L. Malvaceae) based on sequences of three low-copy nuclear genes. *Plant Syst. Evol.* 252, 199–214.
- Applequist, W. L., Cronn, R. and Wendel, J. F. (2001) Comparative development of fiber in wild and cultivated cotton. *Evol. Development* 3, 3–17
- Beasley, J.O. (1940) The origin of American tetraploid *Gossypium* species. *Amer. Nat.* 74, 285–286.
- Beasley, J.O. (1941) Hybridization, cytology, and polyploidy of *Gossypium*. *Chronica Botanica* 6, 394–395.
- Brubaker, C.L., Bourland, F.M. and Wendel, J.F. (1999) The origin and domestication of cotton. In: C.W. Smith and J.T. Cothren (Eds.) *Cotton; origin, history, technology and production*. John Wiley, New York, pp. 3–31.
- Brubaker, C.L., Koontz, J.A. and Wendel, J.F. (1993) Bidirectional cytoplasmic and nuclear introgression in the New World cottons, *Gossypium barbadense* and *G. hirsutum* (Malvaceae). *Amer. J. Bot.* 80, 1203–1208.
- Brubaker, C.L. and Wendel, J.F. (1993) On the specific status of *Gossypium lanceolatum* Todaro. *Genet. Res. Crop Evol.* 40, 165–170.
- Brubaker, C.L. and Wendel, J.F. (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum* ; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Amer. J. Bot.* 81, 1309–1326.
- Brubaker, C.L. and Wendel, J.F. (2001). RFLP diversity in cotton. In: J.N. Jenkins and S. Saha (Eds), *Genetic improvement of cotton: emerging technologies*. Science Publishers, Inc. Enfield, NH, pp. 81–102.

- Cronn, R.C., Small, R.L., Haselkorn, T. and Wendel, J.F. (2002) Rapid diversification of the cotton genus (*Gossypium* : Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Amer. J. Bot.* 89, 707–725.
- Cronn, R.C., Small, R.L., Haselkorn, T. and Wendel, J.F. (2003) Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. *Evolution* 89, 707–725.
- DeJode, D.R. and Wendel, J.F. (1992) Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *Amer. J. Bot.* 79, 1311–1319.
- Denham, H.J. (1924) The cytology of the cotton plant. II. Chromosome numbers of Old and New World cottons. *Ann. Bot.* 38, 433–438.
- Endrizzi, J.D., Turcotte, E.L. and Kohel, R.J. (1985) Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* 23, 271–375.
- Fryxell, P.A. (1965) Stages in the evolution of *Gossypium*. *Adv. Frontiers Plant Sci.* 10, 31–56.
- Fryxell, P.A. (1968) A redefinition of tribe *Gossypieae*. *Bot. Gazette* 129, 296–308.
- Fryxell, P.A. (1971) Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). *Evolution* 25, 554–562.
- Fryxell, P.A. (1979) *The natural history of the cotton tribe*. Texas A & M University Press, College Station, TX.
- Fryxell, P.A. (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheede* 2, 108–165.
- Galau, G.A. and Wilkins, T.A. (1989) Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident A cytoplasm with that of D species *G. harknessii*. *Theor. Appl. Genet.* 78, 23–30.
- Hamrick, J.L. and Godt, M.J.W. (1990) Allozyme diversity in plant species. In: A.H.D. Brown, M. T. Clegg, A.L. Kahler and B.S. Weir (Eds), *Plant Population Genetics, Breeding and Genetic Resources*. Sinauer, Sunderland, MA, pp.43–63.
- Harland, S.C. (1940) New allopolyploids in cotton by the use of colchicine. *Trop. Agric.* 17, 53–55.
- Hawkins, J.S., Pleasants, J. and Wendel, J.F. (2005) Identification of AFLP markers that discriminate between cultivated cotton and the Hawaiian Island endemic, *Gossypium tomentosum* Nuttall ex Seeman. *Genet. Res. Crop Evol.* 52, 1069–1078.
- Hendrix, B. and Stewart, J.McD. (2005) Estimation of the nuclear DNA content of *Gossypium* species. *Ann. Bot.* 95, 789–797.
- Hutchinson, J.B. (1951) Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* 5, 161–193.
- Hutchinson, J.B. (1954) New evidence on the origin of the Old World cottons. *Heredity* 8, 225–241.
- Hutchinson, J.B. (1959) *The application of genetics to cotton improvement*. Cambridge University Press, Cambridge.
- Hutchinson, J.B. (1962) The history and relationships of the world's cottons. *Endeavour* 21, 5–15.
- Hutchinson, J.B. and Ghose, R.L.M. (1937) The classification of the cottons of Asia and Africa. *Indian J. Agric. Sci.* 7, 233–257.
- Hutchinson, J.B., Silow, R.A. and Stephens, S.G. (1947) *The evolution of Gossypium and the differentiation of the cultivated cottons*. Oxford University Press, London.
- Hutchinson, J. G. and Stephens, S.G. (1945) The seed hairs of *Gossypium*. *Annals Bot.* 9, 361–367.
- Iqbal, J., Reddy, O.U.K., El-Zik, K.M. and Pepper, A.E. (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* 103, 547–554.
- Khan, S.A., Hussain, D., Askari, E., Stewart, J.McD., Malik, K.A. and Zafar, Y. (2000) Molecular phylogeny of *Gossypium* species by DNA fingerprinting. *Theor. Appl. Genet.* 101, 931–938.
- Kim, H. J. and Triplett, B.A. (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiology* 127, 1361–1366.

- Lacape, J.-M., Dessauw, D., Rajab, J., Noyer, J.-L. and Hau, B. (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol. Breeding* 19, 45–58.
- Liu, B., Brubaker, C.L., Mergeai, G., Cronn, R.C. and Wendel, J.F. (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44, 321–330.
- Liu, D., Guo, X.P., Lin, Z., Nie, Y.C. and Zhang, X. (2005) Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. *Genet. Resources Crop Evol.* 53, 1145–1152.
- Liu, Q., Brubaker, C.L., Green, A.G., Marshall, D.R., Sharp, P.J. and Singh, S.P. (2001) Evolution of the *FAD2-1* fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *Amer. J. Bot.* 88 92–102.
- Liu, S., Cantrell, R.G., McCarty, J.C.J. and Stewart, J.McD. (2000) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *Crop Sci.* 40, 1459–1469.
- McCarty, J.C. and Percy, R.G. (2001) Genes from exotic germplasm and their use in cultivar improvement in *Gossypium hirsutum* L. and *G. barbadense* L. In: J.N. Jenkins and S. Saha (Eds.), *Genetic improvement of cotton: emerging technologies*. Science Publishers, Inc. Enfield, NH, pp. 65–80.
- Meredith, W.R. (1991) Contributions of introductions to cotton improvement. In: H.L. Shands and L.E. Wiesner (Eds.), *Use of plant introductions in cultivar development, part 1* *Crop Sci. Soc. Amer.* Madison, pp. 127–146.
- Meyer, V.G. (1974) Interspecific cotton breeding. *Econ. Bot.* 28, 56–60.
- Narayanan, S.S., Singh, J. and Varma, P.K. (1984) Introgressive gene transfer in *Gossypium*: Goals, problems, strategies and achievements. *Cot. Fib. Trop.* 39, 123–135.
- Niles, G.A. and Feaster, C.V. (1984) Breeding. In: R.J. Kohel and C.F. Lewis (Eds.) *Cotton*, *Amer. Soc. Agronomy*, Madison, pp. 201–231.
- Percy, R.G. and Wendel, J.F. (1990) Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor. Appl. Genet.* 79, 529–542.
- Phillips, L.L. (1966) The cytology and phylogenetics of the diploid species of *Gossypium*. *Amer. J. Bot.* 53, 328–335.
- Pillay, M. and Myers, G.O. (1999) Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Sci.* 39, 1881–1886.
- Rapp, R.A., Alvarez, I. and Wendel, J.F. (2005) Molecular confirmation of the position of *Gossypium trifurcatum* Vollesen. *Genet. Resources Crop Evol.* 52, 749–753.
- Rungis, D. D., Llewellyn, D., Dennis, E.S. and Lyon, B.R. (2005) Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Aust. J. Agric. Res.* 56, 301–307.
- Ryser, U. and Holloway, P.J. (1985) Ultrastructure and chemistry of soluble and polymeric lipids in cell walls from seed coats and fibres of *Gossypium* species. *Planta* 163, 151–163.
- Saunders, J.H. (1961) *The wild species of Gossypium and their evolutionary history*. Oxford University Press, London.
- Seelanan, T., Brubaker, C.L., Stewart, J.McD. Craven, L.A. and Wendel, J.F. (1999) Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Syst. Bot.* 24, 183–208.
- Seelanan, T., Schnabel, A. and Wendel, J.F. (1997) Congruence and consensus in the cotton tribe. *Syst. Bot.* 22, 259–290.
- Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J.K., Noyes, R.D., Paterson, A.H. Wing, R.A., Wilkins, T.A. and Wendel, J.F. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Biol. Evol.* 20, 633–643.
- Skovsted, A. (1934) Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet.* 28, 407–424.
- Skovsted, A. (1937) Cytological studies in cotton. IV. Chromosome conjugation in interspecific hybrids. *J. Genet.* 34, 5–42.

- Small, R.L. and Wendel, J.F. (1999) The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). *J. Hered.* 90, 251–253.
- Small, R. L. and Wendel, J. F. (2000) Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L. Malvaceae). *Mol. Phyl. Evol.* 16, 73–84
- Stanton, M.A., Stewart, J.McD., Percival, A.E. and Wendel, J.F. (1994) Morphological diversity and relationships in the A-genome cottons, *Gossypium arboreum* and *G. herbaceum*. *Crop Sci.* 34, 519–527.
- Stewart, J.McD. (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: G.A. Constable and N.W. Forrester (Eds.), *Challenging the Future: Proceedings of the World Cotton Research*. CSIRO, Melbourne, pp. 313–327.
- Stewart, J.McD., Craven, L.A., Brubaker, C.L. and Wendel J.F. (2008). *Gossypium anapoides* : a new species of *Gossypium*. Novon (In press)
- Ulloa, M., Brubaker, C.L. and Chee, P. (2007) Cotton. In: C. Kole (Ed.) *Genome Mapping and Molecular Breeding in Plants, Volume 6, Technical Crops*. Springer, Berlin, pp. 1–50.
- Ulloa, M. Stewart, J.McD. Garcia-C. E.A. Godoy-A.S, A. Gaytán-M. and Acosta-N. S. (2006) Cotton genetic resources in the western states of Mexico: *in situ* conservation status and germplasm collection for ex situ preservation. 2005. *Genet. Resources Crop Evol.* 53, 653–668.
- Wang, G.-L., Dong, J.-M. and Paterson, A.H. (1995) The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germplasm: molecular analysis of introgressive plant breeding. *Theor. Appl. Genet.* 91, 1153–1161.
- Watt, G. (1907) *The wild and cultivated cotton plants of the world*. Longmans, Green and Co. London.
- Webber, J.M. (1935) Interspecific hybridization in *Gossypium* and the meiotic behavior of F₁ plants. *J. Agric. Res.* 51, 1047–1070.
- Wendel, J.F. (1989) New World tetraploid cottons contain Old World cytoplasm. *Proc. Natl. Acad. Sci. USA* 86, 4132–4136.
- Wendel, J.F. and Albert, V.A. (1992) Phylogenetics of the cotton genus (*Gossypium* L.): Character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot.* 17, 115–143.
- Wendel, J.F., Brubaker, C.L. and Percival, A.E. (1992) Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *Amer. J. Bot.* 79, 1291–1310.
- Wendel, J.F. and Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186.
- Wendel, J. F. and Percy, R.G. (1990) Allozyme diversity and introgression in the Galapagos endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* 18, 517–528.
- Wendel, J.F., Stewart, J.McD. and Rettig, J.H. (1991) Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium*. *Evolution* 45, 694–711.
- Wendel, J.F., Olson, P.D. and Stewart, J.McD. (1989) Genetic diversity, introgression and independent domestication of Old World cultivated cottons. *Amer. J. Bot.* 76, 1795–1806.
- Wendel, J.F., Rowley, R. and Stewart, J.McD. (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Pl. Syst. Evol.* 192, 49–59.

The Worldwide Gene Pool of *G. hirsutum* and its Improvement

Edward L. Lubbers and Peng W. Chee

Abstract The *Gossypium* genus has more than 50 species that are available to use as germplasm; 5 species, including *G. hirsutum*, are allotetraploids while the rest are diploid species. The 7 races of *G. hirsutum* are directly usable as a germplasm resource with photoperiodism as the main barrier. The sister tetraploids require further effort to be utilized as germplasm for *G. hirsutum* improvement due to segregational breakdown. Utilizing germplasm from the diploids generally requires more extreme methods such as chromosome doubling and the use of bridging species. The phenotypic consequences of the domestication of Upland cotton, *G. hirsutum*, are similar to the domestication syndrome that is generally common to many crop plants. The genetic consequences of domestication are reflected by the very low level of genetic diversity found in Upland cultivars. In continuing the domestication of Upland cotton, developmental breeding programs are quickly becoming absolutely imperative to provide the diversity that is needed to provide intrinsic genetic solutions to the needs of producers, processors, and consumers.

1 Introduction

Cotton from *Gossypium* L. species has been a fiber component of textiles and other manufactured items for more than 5000 years in the New World (Damp and Pearsall 1994; Dillehay, Rossen, Andres, and Williams 2007); and for more than 4300 years in the Old World (Gulati and Turner 1928). Renewable fiber resources from cotton as an alternative to synthetic fibers are becoming more valuable in this age of high crude oil prices. Unfortunately, in most of the world, petroleum resources are also heavily used in the production of cotton. This dichotomy reinforces the need to more efficiently provide greater production and quality which, in turn, increases the value of genetic resources. Transgenic

E.L. Lubbers (✉)

Department of Crop and Soil Sciences, University of Georgia, Tifton, Georgia, USA
e-mail: elubbers@uga.edu

improvement will become more valuable as the knowledge base of domestication grows, but the germplasm resource will become increasingly important not only as a resource but as a foundation of the knowledge base. Even though the main importance for cotton is fiber, the secondary products are also important in adding value to the crop. Cottonseed oil is a source of vegetable oil for culinary uses and a raw material for biofuel. Cottonseed meal is an animal feed and fertilizer.

While *Gossypium* species are endemic to the tropics and subtropics, the 33 million hectares of cotton that are planted annually in almost 70 countries (ICAC 2004), ranging from latitudes 47°N in Ukraine (UNCTAD 2006) to about 34°S in Australia (AOGTR 2002), show the importance of cotton in today's global economy. Annual cotton fiber production has now reached about 25 million metric tons with a farm-gate value of about US\$20 billion, contributing about 40% to the world fiber market (Townsend 2006) and making cotton the single most important natural fiber in the global economy. The value of the processing of cotton adds US\$10 billion to the farm-gate for an aggregate value estimated to be US\$30 billion. More than 350 million people are engaged in jobs related to the production and processing of cotton. Of the *Gossypium* species, *G. hirsutum* L. is the focal point of this chapter since it is the primary source of cotton fiber, accounting for about 97% of the world production (NCCoA 2006).

2 Evolutionary Structure

Understanding the evolutionary structure of the crop is imperative to be able to efficiently and effectively breed for improvement (see Chapter 1 by Wendel, Brubaker, Alvarez, Cronn, and Stewart, this volume). To describe the evolutionary structure of *G. hirsutum*, one can start at the level of the tribe Gossypieae as the furthest likelihood that a reproductive gene pool would be recognizable. *Gossypium* has been found to be monophyletic and shares a most recent ancestor with the genera *Kokia* and *Gossypoides*. The *Gossypium* genus has more than 50 species that are available to use as germplasm; 5 species, including *G. hirsutum*, are allotetraploids while the rest are diploid species. Closer specific relationships of *G. hirsutum* to the germplasm of choice allow an easier exploitation of novel genes and alleles in placing them into commercially valuable germplasm.

2.1 *Gossypium* Gene Pools

Harlan and de Wet (1971) proposed an extension of the definition of the gene pool to emphasize genetic compatibility over the strict taxonomic placement of species. The original idea presupposes utilizing the reproductive apparatus of

the plants to transfer the genes and separates the difficulty to do so into three levels; primary to secondary to tertiary, from easy to difficult.

Stewart (1994) has previously assigned the *Gossypium* genome groups to gene pools. In this categorization, the primary gene pool consists of all the allotetraploid *Gossypium* species with the diploid species split between the secondary and tertiary gene pools based on their relationship with the A and D genome progenitors of those *Gossypium* allotetraploids. This places the A and D genomes with the B and F genomes, having chromosomes that are structurally similar to the A and D genomes, in the secondary gene pool. Once a fertile hybrid is produced, these genomes have a relatively high recombination frequency. The C, E, G, and K genomes are placed in the tertiary pool based on the low level of genetic recombination found once the extremely difficult production of a hybrid is accomplished.

Although this assignment has value, the overall categorization diverges from Harlan and de Wet's format. First, it is appropriate to summarize from Harlan (1992): the primary gene pool corresponds with the species of the crop where hybrids are generally fertile with good chromosome pairing, gene segregation is approximately normal, and gene transfer is simple. The secondary gene pool includes the species that approximate experimentally defined coenospecies (a group of individuals of common evolutionary origin comprising more than one species). The coenospecies is made up of ecospecies (a group of ecotypes capable of interbreeding without loss of fertility or vigor in the offspring) which following hybridization may participate to a limited extent in mutual change of genes. Furthermore, gene exchange between different coenospecies does not occur; sterile hybrids are produced by hybridization (Rieger, Michaelis, and Green 1976). Recovery of desired types in advanced generations may be difficult, but the tertiary gene pool is available with some effort. This group can be crossed but the hybrids are anomalous, lethal, or completely sterile. At best, gene transfer requires extreme measures such as chromosome doubling, embryo culture, grafting or tissue culture, or the use of bridging species.

Therefore, by following the concept outlined by Harlan and de Wet, everything within the species *G. hirsutum* is in the primary gene pool. Next, the description of the secondary gene pool fits the tetraploid *Gossypium* species as a coenospecies almost perfectly. As an example, *G. hirsutum* and *G. barbadense* crosses are vigorous and fertile but these interspecific hybrids show genetic breakdown during segregation in subsequent generations (Hutchinson 1951). This genetic breakdown can be seen as segregation distortion of DNA markers sometimes showing no introgression within a number of chromosomal regions or even entire chromosomes (Jiang, DelMonte, Paterson, Wright, and Woo 2000). Finally, the gene pool of the diploid *Gossypium* species and any more distant groups that are successfully used sexually, would be in the tertiary gene pool since genes from these species must, at present, be acquired by utilizing some of the extreme methods listed above by Harlan (1992) such as chromosome doubling and bridging species. Some researchers have further proposed designating the quaternary gene pool to indicate transgenic manipulation in

transferring specific genes from organisms more distant genetically such as the *Bt* trait in cotton that comes from the bacteria *Bacillus thuringiensis*. Stewart (1995) though designates transgenic manipulation as the $\alpha\omega$ (alpha omega) gene pool to correctly indicate that these genes can literally come from any organism no matter which gene pool it may fit into.

3 Cotton Domestication

A quintessential definition of domestication from a plant breeder's point of view is "a coevolutionary process by which human selection on the phenotypes of . . . plant populations results in changes in the population's genotypes that make them more useful to humans and better adapted to human intervention" (Clement 1999; Paterson 2002). Domestication can range from the first selection of a more desired plant from the wild population to modern cultivars that have been strongly and successfully selected over historical time. Since selection implies the desire to keep the change, domesticating a plant inherently changes the genomic makeup of the retained population both by keeping desirable traits as well as eliminating the undesirable traits. As one can see from this, the original domesticators were the first plant breeders, a role that continues to be of paramount importance in modern civilization.

Domestication is microevolution under the influence of man. All four of the domesticated cotton species probably were domesticated independently by four geographically different civilizations (Fryxell 1979). A probable scenario involves domestication as door-yard agriculture that developed into field production as civilization became more industrialized and agriculture became more specialized. Some traits that may have been selected as cotton was domesticated include reduced seed dormancy to better control plant density, increased yield by controlling the size and number of bolls per plant and lint percentage, reduced plant size to a scale that humans can utilize, and a more annual life cycle. Later selections were made to improve the quality of the fiber, initially fine, stronger fibers that were more uniform and longer. Additional fiber qualities such as elasticity and short fiber content have also become important in the textile industry. As agriculture becomes more agribusiness, the industrialization of production will continue to define traits that increase efficient production. Examples include the increase in mechanization that has required selection for appropriate plant stature and the susceptibility of monoculture to pests and disease that requires increased selection for resistant genotypes.

3.1 *The Races of G. hirsutum*

Because the wild and cultivated forms of a species are genetically interconnected, internationally agreed taxonomic rules dictate that once a wild ancestor is adequately recognized then it and the cultivated form(s) shall be considered a

single species and named accordingly (Zohary and Hopf 2000). It thus follows that the germplasm of domestication begins with the subspecific ranks, the races or varieties. There are 7 races of *G. hirsutum* that are presently recognized (Hutchinson 1951); ‘yucatanense’, ‘punctatum’, ‘palmeri’, ‘latifolium’, ‘marie-galante’, ‘morrilli’, and ‘richmondi’. Of these seven, ‘punctatum’, ‘latifolium’, and ‘marie-galante’ have dispersed the furthest with ‘latifolium’ being considered the race from which the modern cultivated “Upland” cotton was derived. Of these races, only ‘yucatanense’ is considered to exist solely as a wild form; all the others have traits that could be considered valuable in a domesticated form of cotton. During the botanical collection surveys, all races other than ‘yucatanense’ were strongly associated with man and his activities.

3.1.1 Phylogenetic Categorizations

Wendel, Brubaker, and Percival (1992) used allozyme variation to verify the existence of an underlying geographical component of sub-specific genetic relationships among these races, but were unable to strongly support Hutchinson’s racial divisions. Hutchinson (1951) acknowledged that the barriers between the races are not absolute in that it is often impossible to assign a given specimen to its correct race. Brubaker and Wendel (1994) revisited the question of the origin of cotton using DNA variation detected by RFLPs (Restriction Fragment Length Polymorphisms) which led to a clearer picture of the relationships among the races. Whereas Wendel et al. (1992) essentially were able to show a general separation of the specimens from the Caribbean (generally the ‘marie-galante’) from those found on mainland Mesoamerica, Brubaker and Wendel (1994) showed that the accessions fundamentally divided into three groups; (1) ‘marie-galante’, (2) ‘yucatanense’ and ‘punctatum’, and (3) ‘latifolium’ and ‘palmeri’. Interestingly, internal manuscript comments by S. G. Stephens (Hutchinson 1951) show his preference for fewer divisions that would also fit ‘marie-galante’, ‘punctatum’, and ‘latifolium’ as classes. The accessions used by Lacape, Dessauw, Rajab, Noyer, and Hau (2007) showed full support for the racial classification suggested by Hutchinson (1951). As further analyses with greater number of markers are utilized, the intraspecific relationships are being clarified. Even though further research needs to be done since there are irregularities in the work such as all of a particular race coming from a single location, the basic framework shown by Hutchinson (1951) continues to be supported and clarified by subsequent research.

The likely basal lineage of the races of *G. hirsutum* reported by Brubaker and Wendel (1994) was from a group within the ‘yucatanense’ and ‘punctatum’ races which are either wild or minimally domesticated (Hutchinson 1951). The sister lineage of ‘latifolium’ and ‘palmeri’ exists only under domestication and have greater evidence of artificial selection. Lacape et al., (2007) supports this racial separation and goes further to separate ‘yucatanense’ from ‘punctatum’. This leads to a geographical premise that the coastal Yucatan peninsula is the primary location of cotton domestication with the primitive domesticated

landrace 'punctatum' being derived from 'yucatanense' with further branching occurring from 'punctatum' to 'latifolium', 'palmeri', 'richmondi', and 'morrilli' occurring more inland (Brubaker and Wendel 1994; Lacape et al., 2007). The race 'latifolium' is found more specifically in southern Mexico/Guatemala. Morphologically this is supported by 'punctatum' being intermediate between the wildest forms and the more manipulated forms like 'palmeri' and 'latifolium' (Hutchinson 1951). Since race 'yucatanense' has the characteristics of truly wild species (specifically small seeds with hard seed coats, delayed germination, and small bolls with brown, short and scanty lint) and morphologically intergrades with 'punctatum', this supports the idea that 'yucatanense' is the wild *G. hirsutum*.

Brubaker and Wendel (1994) considered the remaining race, 'marie-galante' of Central America, a phylogenetically and phenetically isolated race with little gene flow between these accessions and the lineages of the other races they studied. It is the most tree-like of all the cottons (Hutchinson 1951) with a key characteristic being strong apical dominance of the main stem. This is illustrated by 10 private alleles, i.e. found in neither the other *G. hirsutum* nor the *G. barbadense* accessions. As are all the other races except 'latifolium', 'marie-galante' is considered a perennial. It is found from Guatemala to Colombia to Ecuador, in the Caribbean, and in northern South America in areas sympatric with *G. barbadense*. This in turn led Stephens (1967) to propose that 'marie-galante' arose via introgression between *G. hirsutum* and *G. barbadense*. Even though there is sufficient data to show that 'marie-galante' has a history of *G. barbadense* alleles being introgressed into *G. hirsutum* (Stephens 1967; Stephens and Phillips 1972; Wendel et al. 1992; Brubaker and Wendel 1994; Lacape et al. 2007), there is no support that the origin of 'marie-galante' is interspecific. There is, however, good evidence for introgression from *G. barbadense* and the genetic distance from *G. barbadense* to 'marie-galante' is greater than 'yucatanense' to *G. barbadense* (Lacape et al., 2007), thus placing 'marie-galante' solidly in the *G. hirsutum* clade. This leads to one possible interpretation that 'marie-galante' was a somewhat domesticated *G. hirsutum* that spread across the Caribbean and northern South America where it was pollinated by the sympatric *G. barbadense*. This would, of course, decrease the genetic distance from *G. barbadense* when measured in the present. Another possibility may be separate domestication event for 'marie-galante' (Stephens 1972) in Colombia. No apparent transition zone is found between 'marie-galante' and the other *G. hirsutum* races in Central America. Also, the relative morphological uniformity of 'marie-galante' along the Pacific coast of Central America compared to Colombia is opposite of what would be expected if there was a common center of domestication in southern Mexico and Guatemala. Stephens (1972) presented a distribution of accessions with unique seed, boll, bracteole, and leaf traits that can be easily explained by following the topography out of this area of Colombia. Even though he considered this another center of domestication, there were no wild populations found along the Colombian coastline which is where wild forms of cotton are usually found.

Since present day Upland cotton is most closely related to 'latifolium', enhancing the genetic diversity of Upland cotton with 'latifolium' germplasm would have the fewest barriers. The accessions gathered from southern Mexico/Guatemala are highly variable in plant habit, hairiness, anthocyanin pigmentation, boll size, productivity, etc. (Hutchinson 1951). The remaining races, 'palmeri', 'richmondi', and 'morrilli', are not spread as widely as 'punctatum', 'marie-galante', and 'latifolium' and have not been researched in as much detail but it is clear that they are closely related to the race 'latifolium' and the Upland cultivars (Brubaker and Wendel 1994; Lacape et al. 2007). They should be considered as excellent resources within the primary gene pool for increasing the diversity of the Upland cotton germplasm. The hybrids of any of these races are fertile and the frequency of recombination is high within the entire primary gene pool.

3.2 Upland Cotton

Cotton as a product was introduced into the eastern coastal areas of North America as the Europeans immigrated into this area. All fiber producing species of *Gossypium* were tried as crops early in the history of the United States. Both New World tetraploids were found to produce better than the Old World diploids. *G. hirsutum* was grown in the upland areas of the country whence it received its common name of Upland cotton. Most of the improvement of *G. hirsutum* as Upland cotton arose within the Cotton Belt of the USA (Ware 1951). Although Upland cotton genetics is a product of the 'melting pot' of imported cotton, the material from the Mexican highlands that came into the cotton belt in the first part of the 19th century via Texas is certainly a foremost source of the current genetic structure of modern Upland cotton. Ware (1951) gave an account of the history of the development of Upland cotton cultivars from the time of the immigration of European settlers to the middle of the 20th century. The initial thrust of this emphasis on improved cotton production was the use of large-scale mechanization in England to produce cotton textiles.

3.2.1 United States Cotton Belt Groupings

As cotton became a commercialized enterprise beyond a cottage industry, distinctive cultivars arose with adaptation to specific local environments. These distinctions lead to the formation of types but never to the level of botanical differences found in sub-species or species. Between Duggar (1907) and Tyler (1910), nine types were described at the height of differentiation of Upland cultivars which was just before the arrival of the boll weevil. The groups were the Eastern Big Boll type, Western Big Boll type, Semi-cluster type, Cluster type, Rio Grande type, Early type, Long Limb type, Upland Long Staple type, and the Intermediate or Miscellaneous type. The following descriptions are to

illustrate the phenotypic diversity available in the middle to late 1800s prior to the era of the boll weevil. A great loss of germplasm occurred at this time since most of the cultivars in these groups are no longer available.

The Eastern Big Boll type used the Truitt cultivar as the “type variety”, the cultivar used to describe a typical cultivar in the group. It had 3 to 4 large basal branches with comparatively short branches on the main stalk. The bolls were very large, egg-shaped with a blunt point, and easy to pick as they opened wide. The seeds were gray and fuzzy with a seed index of about 14 grams. The fiber was about 2 to 2.5 cm long with ‘good’ quality. This description and those following are from Ware (1951) which was generally taken from Duggar (1907) and Tyler (1910). The Western Big Boll type had the large, pendant boll size of the Eastern Big Boll type; but they also had bract and bur structures that protected the boll from storm loss along with a greater lock adherence than the cultivars from the Eastern Big Boll type. The type variety for the Western Big Boll was a cultivar called Stormproof (or at times Schuback). Another difference was that the Eastern Big Boll type almost certainly was traceable to a cultivar Wyche that came to the United States by way of Algeria, whereas the Western Big Boll type is considered to have arisen in Texas via importation from the Mexican highlands. Much more of the breeding history of the modern cultivars exploiting the Western Big Boll type has been expounded by Ware (1951).

The Semi-cluster type used Peerless as the type variety. It was moderately tall, shaped like a pyramid with 1 to 3 monopodial branches. The sympodial branches had irregularly spaced nodes and were from 45 cm long at the bottom of the plant to 5 cm at the top. The bolls were small to medium with the lint from 2 to 2.5 cm long. The seeds were fuzzy, greenish or brownish gray, and small. The Cluster type variety was Dickson. The Cluster type differed from the Semi-cluster type only in the degree of the cluster plant habit. The sympodial branches were reduced to spurs by the pronounced shortening of the internodes. They were generally longer in the middle of the plant than at the bottom or the top. Both of these types were associated with early maturity but Ware (1951) points out that this may be coincidental.

Peterkin was the type variety for the Rio Grande type. The main traits associated with this type were high lint percentage and very small seeds with sparse to almost no fuzz. There was always a tuft of fuzz on the hilar end of the seed. The plants were slender in growth habit which included long, spindly monopodial branches with slender, drooping sympodial branches. The branches were also reddish from anthocyanin but not intensely. The plants were open branching with long internodes and thin foliage. The bolls were ovate but with pointed tips. Although they opened well, they retained the locks better than most bolls that opened widely. The fiber was considered wiry and strong with length of about 2 cm. The little fuzz that was present was short and brownish gray. The cultivars of this group were unsurpassed for poor, droughty, badly maintained hill land, and the associated difficult farming conditions.

The Early type cultivars were defined by the extreme earliness of their growth habit. The type variety was King. Two cultivars, Maryland Green Seed and Tennessee Green Seed, are considered to be related and in this group. They were considered to be even earlier than King. Ware (1951) considered these to be remnant of the cultivars that allowed cotton to be grown as far north as eastern Pennsylvania, New Jersey, Delaware, and Maryland during the American Revolution. Many of the cultivars from this group were considered to be the same when planted in comparative trials. King was slender with 1 to 3 monopodial branches. The sympodial branches had short internodes but did not seem to tend to semi-cluster habit. The leaves were described as medium to small with deep lobes. The creamy white flowers generally had petal spots, uncommon for Upland cotton. The bolls were small and four-locked. The lint was short; the seeds were small with short brownish gray fuzz.

The Long Limb type, as the name implies, is very large with long monopodial and sympodial branches. They are slender with long internodes and tend to hang down when loaded with bolls. The growth habit is open and late maturing. The leaves were medium-sized and the medium bolls were round to ovate with blunt points. The seeds mostly had fuzz that was greenish to whitish brown color and lint that was not considered very long. The type variety, Petit Gulf, was named for the place on the Mississippi River south of Vicksburg, MS from which cotton of this type was shipped. The Upland Long Staple type was similar to the Petit Gulf type except that it had long staple and better lint percentage. It did, however, seem to have a longer pointed boll. Ware (1951) indicated that the parent stocks of the Upland Long Staple were of Mexican origin rather than from any introgression. It was felt that the diversity at the time was adequate to make available the greater length, fineness, and other characters found in the Upland Long Staple. There were a number of claims for cotton produced from the cultivars in this type that would be valuable for breeding today. It is not known how much is hyperbole and how much of this "exquisite" quality that modern breeders were actually able to capture in the present cultivars. The type variety was Allen but the descriptions of other cultivars in this group were more diverse than for other types. It is obvious that fiber quality was an important phenotype in the early breeding efforts of Upland cotton that is being maintained today. Both the Upland Long Staple and the Western Big Boll types are pivotal in the development and history of cotton breeding in the United States cotton belt.

At the beginning of the boll weevil invasion, the United States Department of Agriculture led an attempt to significantly modify cultivars in order to allow growers to produce cotton in the presence of this devastating insect utilizing plant-to-row breeding techniques (Ware 1951). As Brownsville, TX was the first place in the United States to have the boll weevil in 1892, Texas was the location where the plans to combat the invasion were formulated and enacted (Brown and Ware 1958). The dispersion of the boll weevil across the United States cotton belt took about 30 years, reaching Virginia around 1922. The southwestern United States has escaped the invasion of the boll weevil. Many of the

cultivars that either survived the boll weevil or were developed for combating the insect came from the Western Big Boll type. From this type, a new series of types was developed which included Mebane Triumph type, Rowden type, Lone Star type, Stoneville type, Coker 100 type, and Deltapine type. Of the important cotton seed companies based on these types, three went on to be key players in the seed cotton industry; Stoneville Pedigreed Seed Co., Coker's Pedigreed Seed Co., and Delta and Pine Land Co. Further improvements continued as any and all of the types were utilized as parents without much consideration for any maintenance of types, thus blending the types.

Additional introductions from Guatemala and Mexico were made to identify genetic diversity that could be used against the boll weevil; Acala, Durango, Tuxtla, and Kekchi (Ware 1951). Acala was expanded into an extensive population and demonstrated a considerable ability to change form, response to the environment, and lint quality. With a boost from the California one-variety cotton law, Acala became synonymous with excellent fiber quality cotton; behind, of course, *G. barbadense* Pima/Egyptian cottons. Durango became popular in the Imperial Valley, CA and became the first cultivar to be selected for a one-variety community. It was eventually replaced by the continually improved Acalas. Tuxtla required a growing season too long for boll weevil conditions. Even though it had potential in the irrigated Southwest where there were no boll weevils, it was not able to compete with either Acala or Durango. Kekchi was selected as small, compact, early and productive in Guatemala but its growth habit in different environments was surprising during the extensive acclimation that was required to develop commercial cultivars. In Texas it was extremely tall (~ 3m) and was sterile for many of the plants while in southern California, it was less vigorous. In Kansas and Maryland it was nearly normal compared to what was reported in Guatemala. Paymaster was one of the cultivars that were finally developed from it.

As the 20th century continued, a new grouping scenario in the United States emerged; the Acala type, the Plains type, the Delta type, and the Eastern type (Niles and Feaster 1984). The Acala type is further grouped into the cultivars produced in primarily in California versus those grown in New Mexico. The Plains type is also named the Stripper type for the equipment used to harvest these storm-resistant to stormproof types developed mostly for the High Plains of the Cotton Belt and surrounding area. Other important characteristics of this type are a comparatively compact growth habit and relatively determinant reproduction that are needed to adapt to the expected drought and cooler temperatures that limit the growing season in this area. The Delta type is essentially a combination of the Stoneville type and the Deltapine type from the previous set of groupings. Deltapine and Stoneville both utilized material from the Western Big Boll (Stormproof) type that was surviving the boll weevil influx with Deltapine further using material from the Upland Long Staple (Allen) type. This has been a very successful type over much of the United States as well as in many areas around the world. The Eastern type also utilized much of the same foundational material as the Delta type. There were also

possible introgressions from the pre-boll weevil lines from this area. Recall that the eastern United States has the longest period of cotton production, which included both Sea Island and Upland cotton, and thus was considered to have the greatest likelihood of wide diversity across the United States Cotton Belt.

3.2.2 Other Domesticated Subtypes

There is another group of *G. hirsutum* in Brazil called mocó cotton that is considered to be a subgroup of ‘marie-galante’. It was at one time believed to have been domesticated from *G. mustelinum* but morphological marker data disproved this (Pickersgill, Barrett, and Andrade-Lima 1975; Freire, Moreira, Santos, and Andrade 1998). There are recognized types of mocó cotton that have been mentioned in the literature and include; mocozinho, mocó-vale-do-serido, francisco-raimundo, cream fiber, genetically improved mocó, mocó resembling *G. barbadense*, and mocó with a tendency to *G. mustelinum* (Freire et al., 1998). Freire et al., (1998), using morphological indices, supported the popular understanding of the introgression of *G. barbadense* and *G. mustelinum* into mocó cotton which can be seen from the names of several of the types. Using a phylogenetic tree, Lacape et al. (2007) showed an indication of the wide diversity of mocó cottons with one mocó cotton accession grouped within ‘marie-galante’, another between ‘punctatum’ and ‘richmondi’ and a third between ‘morrilli’ and ‘richmondi’. The mocó germplasm resource obviously can be beneficial to breeders of more annualized cottons as the difficulties associated with photoperiod and perennial growth habit of the mocó cottons are overcome.

There are two other groups of germplasm that have been derived from ‘latifolium’ and may have useful traits not found in the Upland cotton developed in southern United States. The “Cambodias” are now found primarily in south India (Main 1912) but are also known in Indo-China and the Philippines. It came to India via Cambodia and other parts of Southeast Asia by way of direct introductions to the Philippines from Mexico by the Spanish (Lewton 1925). The second group of ‘latifolium’ cotton is in Africa and came from the Uplands of the United States. The success of these subtypes (Hutchinson 1951) was mainly based on genes for resistance to jassids (*Amrasca biguttula biguttula* (Ishida)) and blackarm (*Xanthomonas campestris* pv. *malvacearum* (Smith) Dye [Syn. - *Xanthomonas axonopodis* pv. *malvacearum* (Smith) Vauterin et al.]). The jassids (a leafhopper) and blackarm (bacterial blight) are as much a threat in Africa as in India and Southeast Asia. Knight and Hutchinson (1950) have shown that resistance to blackarm is found in ‘punctatum’ in the Bahamas and is a source of the blackarm resistance found in the African ‘latifolium’. Since the “Cambodias” went through a different bottleneck than the Uplands and the ‘latifolium’ sent to Africa was crossed with ‘punctatum’, there are likely to be alternate genes or alleles that are present which in turn may be valuable to the Upland cotton breeders throughout the world.

4 Phenotypic Consequences of Cotton Domestication

Plant domestication progresses through four basic stages (Harris 1989): procuring wild plants, producing wild plants, methodically cultivating wild plants, and agriculture based on domesticated plants that are more productive while being dependent on humans for survival. In this way plant domestication/plant breeding can be considered the stable genetic modification of a population of plants that better meets the needs of the human(s) that did the choosing. Even the first step of bringing plants in from the wild to grow nearer to where the first domesticators lived is artificial selection, be it conscious or unconscious selection. Saving the plant parts, like sections of stem to propagate cassava, or seeds to replant is an important step in domestication, but the key innovation was to consciously save the seeds or plant parts from the better plants to replant.

One important consequence of plant domestication is that the phenotypes of the domesticated types are different from the ancestral population from which they were derived. The collection of traits associated with domestication or ‘domestication syndrome’ (Hammer 1984 from Doebley, Gaut, and Smith 2006) includes larger fruits, seeds, and other desired plant parts, changes in plant architecture like apical dominance or determinate growth, and loss of natural seed dispersal. Plants that are likely to be propagated vegetatively as well as those that are perennials can follow different paths to being domesticated than do seed bearing annuals like the cereals (Zohary and Spiegel-Roy 1975; Kislev, Hartmann, and Bar-Yosef 2006). As an example of alternate patterns of domestication, Fuller (2007) shows that seed size appears to be an early trait selected in domesticated grasses whereas pulses and pearl millet are not selected for seed size until later in their development as a domesticate.

Cotton appears to have gone from wild to dooryard types to field production. The wild plants are found on beaches around the Gulf of Mexico, the Caribbean, and perhaps some Polynesian locales. The bolls are small and the lint is sparse. As the name implies, dooryard cottons are found in association with human habitation. From the yards to the fields, cotton was brought under cultivation in an increasingly larger scale. Stephens (1965) categorized *G. hirsutum* samples from Venezuela across the Caribbean to the Bahamas, Central America from southern Mexico to southern Nicaragua supplemented with samples from Hawaii and the Brazilian states of Rio Grande do Norte and Para. Since these samples did not exhibit a continuous distribution according to Stephens, they were “easily” placed into 6 discrete groups; (A) wild, (B) semi-feral, (C) commensal, (D) primitive cultigens, (E) improved native types, and (F) commercial introductions of American Upland cotton. These categories are considered to be associated with increasing levels of domestication. Given that the sample number was adequate, the discrete nature of the distribution could likely indicate selection bottlenecks from domestication at each subsequent step. The authors do question the B group as a step in domestication as it seems to be implied that it is a regression to a

more feral state from something domesticated. This understanding may actually strengthen any statistical separations for the first putative step of domestication by considering the step to be from A to C instead A to B. However, viewing the five groups of B to F as progressively greater effects of human manipulation would still be valid.

4.1 Domestication Syndrome of Cotton

Since cotton was developed from perennials and is not grown primarily for food, the pattern of cotton domestication may not follow the patterns found in other crops. Many traits of cotton domestication can be established by noting the contrasting forms required in the cultivated and natural environments (Stephens 1958). The first of the contrasting forms was in the dispersal of seed; efficient dispersal is needed in the wild but is undesirable in agricultural production. The components of seed dispersal include two traits that appear adversarial; bolls must be open enough to allow easy removal of the locks of seedcotton but the locks cannot be allowed to fall out prior to picking (Stephens 1958). Apparently adversarial positions may not truly be antagonistic. At times when severe weather occurs before harvest, repeatable differences in retention of the locks are noticeable between lines in a breeding program while maintaining an open boll suitable for picking. Many breeders take advantage of these instances to further select for enhanced stormproofness. While these somewhat adversarial traits must be balanced, the environmental norms also must be taken into account. A harvesting method called stripping was developed along with cultivars targeted for the High Plains in west Texas which have to resist high winds. These cultivars are developed with bolls that retain the lint better than a standard ‘picker’ cultivar so the lint will not be blown out of the boll. One cultivar with the locks affixed to the burs “by special attachments at the terminal end of the lint hairs” was named Macha after H.A. Macha from Tahoka, TX who selected it out of another cultivar (Ware 1951). Mechanical strippers remove the entire boll and mechanically clean the lint before it is delivered to the gins where there is further cleaning. The lint is not quite as clean as hand or machine picking and some fiber quality is lost in the cleaning, but this difficulty becomes less important as machinery and agronomic procedures improve.

Another aspect of seed dispersal is the ability of the lint to fluff. A cotton fiber is from an elongated cell on the seed surface that has helical layers of cellulose in the cell wall. As the cell matures and dies, the elongated cell wall collapses to form a convoluted ribbon (Lee 1984). The ability of modern cottons to be held in the boll is associated with this flat, convoluted ribbon-like structure. Unless the seedcotton expands to fill the capsule, the lock of cotton will often fall out (Stephens 1958). If the lint fluffs normally, the seeds and the lint may string out but it generally is held at the base of the boll and will not fall out.

Also, the flattening and convolution give rise to the adhesion of the fibers as they are twisted together in spinning which is a key requirement for manual spinning of fiber (Stephens 1958) and would be an important trait for textile manufacture. This may have been of less importance for the early domestication if selection was for needs other than fabricating textiles, such as use of leaves in medicinal infusions for expelling intestinal worms, for lint to stuff pillows and bedding, for plugging cavities in teeth, for packing minor wounds, for wicks for lamps, etc. (Stephens 1958). This is an example of the need to know what was selected first to develop a scenario for the course of domestication. If cotton was used largely for other purposes than making textiles, then traits that are essential for textiles but are very antithetical to survival in the wild could have been more slowly developed under human direction. It is difficult, at times, to understand how a plant could survive in the wild with traits that are desirable in cultivation if they provide a major change in survivability between the cultivated state and the wild state. On the other hand, there may be a flawed understanding of what are truly opposing traits.

In moving from dooryard cottons to small-scale field agriculture, vigorous seedlings and reliable germination were needed (Fryxell 1979) which are generally found by using larger seeds (that have more energy resources for emergence) and which lack hard seed coats. In comparing the progression of domestication, hard seed coats were found to be frequent in the wild group and became less common at every step of domestication, along with an increase in seed size (Stephens 1965). A hard seed coat inhibits the imbibition of water during emergence, keeping the seed from germinating immediately. In the wild, it is a desirable trait; in production agriculture, it is inefficient in developing a crop stand and an otherwise desirable plant may later be a weed in an alternate crop in later growing seasons. Prompt germinability became critically important only as annual cropping necessitated early and uniform emergence to mature the cotton crop in one season (Stephens 1958). Hard seed coat is found in many of the perennial cultivated or wild cottons and is considered the cause of the variability of germination over weeks and months, and in the wild perhaps even years. Modern cotton cultivars have lost hardseededness, but there is an apparent seed dormancy in some breeding lines that return from counter-seasonal breeding nurseries that keeps them from being fully utilized by the breeding companies (Lubbers unpublished). Air-drying at about 40 C for 2 days or with a hot water treatment usually suffices to allow these lines to germinate promptly.

Although the lint percentage of cultivated cottons generally range from over 40% down to about 25%; when adding the putative wild cottons, this range is extended to another peak found at 8 to 10%, thus giving a bimodal distribution (Fryxell 1979). Also, in comparing the amount of lint per 100 seeds, a very highly significant three fold increase occurs between wild and semi-feral samples (Stephens 1965). These two pieces of evidence show that the percentage of lint in seedcotton was selected early in the domestication process of *G. hirsutum* and may have been one of the first traits to be selected. A higher lint percentage

is now found to be correlated to smaller seeds and greater lint yield in modern breeding programs (Miller and Rawlings 1967; Bridge, Meredith, and Chism 1971; Stewart and Kerr 1974). This has led to some breeding lines and possibly some cultivars having difficulty in emerging in stressful environments such as cool soils or through hard soil crusts. But, Culp and Harrell (1975) were able to increase lint yield with only minor increases in lint percentage in their program. They believed some other unknown factors were responsible for their major yield improvements.

Fiber length has also increased during domestication from wild to commensal to primitive lines to advanced types to present cultivars. Stephens (1965) saw a 22% increase from the wild to the modern cultivars. Applequist, Cronn and Wendel (2001) found that in comparing a modern Upland cultivar with an example of 'yucatanense', the fiber length obviously was increased due to artificial selection. This increase was caused by prolonging the total duration of the fiber elongation phase as well as earlier initiation of the maximum rate of post-anthesis fiber elongation. A trait that describes the uniformity of the length of fiber, the uniformity ratio ([the mean fiber length/upper half mean fiber length] times 100), was reported as not being affected by domestication (Stephens 1965).

Selection for larger bolls (larger fruits are another trait in the domestication syndrome) came later since many of the dooryard cottons have bolls of the same size as the wild cottons growing near the seashores. The cottons that were placed in field production have comparatively large bolls (Fryxell 1979) which would assist in efficient hand picking. As cotton production has further intensified in the latter 20th and early 21st century, emphasis is being placed on yield per unit land area which necessitates a balance between number and size of the bolls. The development and use of mechanical cotton pickers limits the selection pressure for larger bolls that are at least unconsciously selected because they are easier to pick by hand.

Annuals are only known in the cultivated cotton species. Both Fryxell (1965) and Stephens (1976) considered the development of a day-neutral reproductive habit, allowing cotton to be grown in temperate as well as tropical environments, to be an essential, major domestication step that was directed by artificial selection. This became important as cotton became a valuable commodity, traded and grown in areas distant from its natural habitat. Of the primitive domesticated *G. hirsutum* material, only 'latifolium' has members that are day-length neutral (Stephens 1976). This leaves the valuable germplasm outside of 'latifolium' photoperiodic and a primary barrier to effectively introgressing desirable traits into a breeding program.

Genetic studies of these traits that were selected during the domestication of Upland cotton are needed. Humankind now has the tools that will enable us to begin to engineer plants for those products that our civilization requires (Vaughan, Balazs, and Heslop-Harrison 2007). An emphasis on research to use crosses/comparisons between different domesticated stages of cotton and truly wild *G. hirsutum*, not feral revertants (Pickersgill 1977), is clearly required

to illuminate the domestication process. A knowledge base is essential to better understand what was selected by the early plant breeders and how these selections impacted the genome so we can better direct further manipulations to enhance production.

5 The Genetic Consequences of Domestication

The first obvious genetic change caused by domestication is in the genes that control the specific phenotypes selected by the original domesticators. Identification of these alleles is just beginning; some classic examples include *teosinte branched1* (Doebley 2004) and *teosinte glume architecture1* (Wang, Nussbaum-Wagler, Li, Zhao, Vigouroux, Faller, Bomblies, Lukens, and Doebley 2005) in maize, *fruitweight2.2* (Frary, Nesbitt, Frary, Grandillo, van der Knaap, Cong, Liu, Meller, Elber, Alpert, and Tanksley 2000) in tomato, *Q* (Simons, Fellers, Trick, Zhang, Tai, Gill, and Faris 2006) in wheat, and *shattering4* (Li, Zhou, and Sang 2006) and *qSH1* (Konishi, Izawa, Lin, Ebana, Fukuta, Sasaki, and Yano 2006) in rice. Other genes that have been shown to differ between cultivars or are important agronomic traits (reviewed in Doebley et al. 2006) can easily be considered as steps in the domestication process. Current QTL research in cotton is likely to uncover some of the specific genes important in the early domestication of cotton.

5.1 Genetic Similarity of the Modern Germplasm

Selection for specific phenotypes affects more than the individual genes underlying those phenotypes. A bottleneck caused by selection will reduce all of the available genetic diversity of a population (Wright 1931; Nei, Maruyama, and Chakraborty 1975). Studying the levels and patterns of genetic diversity in comparing appropriate populations can provide insight into domestication processes. Knowledge of pedigrees gives a beginning to an understanding of the genetic diversity among cotton lines. May, Bowman, and Calhoun (1995) utilized a cluster analysis of the coefficient of parentage (Malécot 1948 from May, Bowman, and Calhoun 1995) to determine the diversity structure among 126 Upland cotton cultivars released between 1980 and 1990. The pedigrees were developed by Calhoun, Bowman, and May (1994) with the coefficients of parentage calculated by employing assumptions appropriate to self-pollinated crops by Bowman, May, and Calhoun (1997). The resulting mean coefficient of parentage of 0.07 implied that these cultivars are genetically diverse, i.e. are not closely related. However, the accuracy of this analysis depends on the accessibility and the accuracy of the pedigrees and on the legitimacy of the analytical assumptions that are made.

The relationships of any historical introductions among themselves and/or between themselves and the extant cultivars of the time are almost always unknown, and a standard assumption is that ancestral introductions are unrelated. Van Esbroeck, Bowman, May, and Calhoun (1999) compared the coefficients of parentage to a genetic similarity index from agronomic and fiber properties for two current cultivars and twelve distant parental cultivars. They found little agreement between the two measures of similarity; the index measure showed a narrow genetic base in cotton. Since the ancestral cultivars were highly similar, the above assumption for the coefficient of parentage was modified to use the mean agronomic/fiber properties index (0.38) as the degree of relatedness of the ancestral cultivars instead of equaling zero. This did not change the relative indications of the relationships but it did increase the coefficient of parentage (from 0.16 to 0.46) and thus decreased the estimation of the genetic diversity developed from previous research (Van Esbroeck, Bowman, Calhoun, and May 1998).

DNA molecular marker technology now provides for analysis of genetic relationships in an absence of pedigree knowledge and has been used within the cotton tribe as stated earlier (Brubaker and Wendel 1994; Tatineni, Davis, and Cantrell 1996; Pillay and Myers 1999; Iqbal, Pepper, El-Zik, and Reddy 2001). All of these corroborated a narrow genetic base of *G. hirsutum* but do not explain the apparent diversity that was reported in cultivars of the late 19th century (Duggar 1907; Tyler 1910) and the moderate diversity in isozymes reported by Wendel et al. (1992). One would not expect phenotypic or isozymatic diversity to be greater than the foundational diversity of DNA. These unexpected findings require further investigation. In comparing DNA marker-based genetic distance measures to the pedigree-based distance, there is only a moderate correlation (Van Becelaere, Lubbers, Paterson, and Chee 2005). This is not unexpected in that DNA marker diversity estimates and the coefficient of parentage estimates differ as to the type of genetic resemblance measured; alike in state vs. identical by descent, respectively.

The lack of genetic diversity is explained by the bottlenecks that are obvious in the development of a crop and this is true of Upland cotton; these bottlenecks include the origin of tetraploid cotton, the initial domestication event(s), and the subsequent movements to new production areas and breeding efforts. A discontinuous distribution in the morphology of *G. hirsutum* (Stephens 1965) implies likely bottlenecks at each major alteration from strong selection pressure. Although the introductions from Mexico to the United States have been considered to have been a potential bottleneck, Iqbal et al., (2001) found that Acala and Tuxtla introductions from Mexico and Kekchi from Guatemala showed no greater genetic diversity than the samples in their collection of Upland cultivars that were mostly from the United States. The Moencopi cotton from the Hopi (native Americans of northeast Arizona) was found to be in the same systematic grouping as the Acala, Tuxtla, and Kekchi introductions and indicates that this type of Upland cotton was obviously widely traded (Iqbal et al. 2001). In turn, this indicates that further enhanced variability will

be more likely found outside of the lines of 'latifolium' that come from the Mexican highland.

Since genetic variation is essential in breeding and introgression with more distantly related material usually brings undesirable effects, crossing adapted germplasm with other groups of cultivated germplasm would utilize the variation between the groups without the introduction of large amounts of undesirable traits. The eastern region of the United States comprises the southeastern and mid-south subregions which include some of the oldest United States cotton production areas. This area includes several important commercial cotton breeding programs as well as some highly regarded public breeding programs. Lubbers, Chee, May, Gannaway, and Paterson (2005) sampled 115 varieties and germplasm lines from the early 1890s to the 1990s. Measures of gene diversity and distance that were calculated using 261 genetically-mapped RFLP markers show a very narrow germplasm base with evidence of previous introgression that was not fully utilized. Little genetic differentiation was found between germplasm from the mid-south versus southeastern subregions with evidence of strong reticulation in the commercial breeding programs of the era. Cotton breeders from these two regions apparently have been using the genetic material from their neighboring region as a germplasm resource. It now looks necessary that the breeding programs will have to range further within the germplasm pool to gain the variability needed to continue to make improved cultivars for the cotton producers. Gaining a better understanding of the genetic relationships among cotton cultivars and germplasm lines will continue to enhance the cotton industry's awareness of the need for further efforts to broaden the genetic diversity available within commercially acceptable genotypes.

6 Cotton Improvement

The early history of the United States provides examples of improved cultivars that were selected out of the production fields, increased, and offered for sale (Ware 1951). Imported germplasm was being constantly tested and selected for improved performance by the cotton planters. The impetus for this effort was the invention of mechanical ginning, spinning, and weaving that made England the dominant player in the textile industry at that time. The introduction of machinery made it necessary for the cotton supply to be larger and more reliable as well as requiring enhanced and more uniform fiber qualities such as fiber length and strength. Presently, new spinning methods are demanding further improvements of fiber quality. A determinate growth habit was also developed to synchronize reproduction and to standardize plant size as mechanized agriculture was brought into the fields. At first, these new requisites were provided by some of the farmers themselves as self-appointed breeders by selecting for increased productivity and quality.

Now the farmers are relying almost completely on the cotton seed companies. For these breeders, two divisions are found within their breeding programs; (1) breeding directly for cultivars that can be competitive in the marketplace and (2) breeding to concentrate desirable phenotypes and remove objectionable traits in order to develop parental material for direct breeding. The latter prebreeding or developmental breeding (Niles and Feaster 1984), is the bridge between genetics and applied breeding. This bridge is absolutely necessary in current breeding programs as it is implausible that any introduction outside of current improved lines/cultivars would be directly used as a cultivar in today's production arena. A successful long-term breeding program requires both applied and developmental components.

6.1 Efforts by the Early Seedsmen

As cotton is considered an inbreeding plant with occasional crossing via natural insect vectors, a simple approach to crop improvement is merely to select from within the variable populations that are currently available. This approach was commonly used by the farmer/seedsman in the early history of cotton production in the United States. With this approach and the importation of diverse cotton germplasm, it appears that high levels of genetic diversity were available early because of the numerous lines that have been described during the pre-boll weevil era (Duggar 1907; Tyler 1910). The numbers of these cultivars may have been inflated by a standard marketing ploy to develop interest in cultivars that were not much different by simply giving them a new name. Nonetheless, the number of cultivars that certainly came from the Western Big Boll type that were foundational for so many of the seed companies in the 20th century (Ware 1951) is remarkable, particularly the numerous selections in a series starting with the cultivar Lone Star that were commercially significant for Stoneville Pedigreed Seed Co. Both mass selection and then plant-to-row methods were used in early efforts in intra-cultivar selection to produce the seed needed to assess the value of the selection criteria of the breeding effort (Ware 1951).

If the necessary variability is not available in the extant populations for selection, then developing that variability by means of crossing is the next step. The importance of selecting desirable parents for crossing was known before the rediscovery of Mendel's work; "Always cross with the seedlings which inherit in the greatest degree the properties you wish a cereal to possess" (Shirreff 1873 from Briggs and Knowles 1967). As Mendel's work was rediscovered and studied, planned hybridizations were established as an integral component to improve the efficiency of the breeding programs. As it was important previously to keep importing and trying new material to discover better cotton lines, it became equally and then more important to find the best parents for the designed crosses. The phenotypes for parents are predicated on what the breeder needs as the combination of traits for the ideal phenotype of the next generation.

6.2 *Contemporary Cotton Breeding*

Applied breeding, the first component in commercial breeding, utilizes either crosses between the more elite cultivars or backcrossing to transfer specific desired traits with the expectation to directly develop a commercial cultivar. As the variability of available elite germplasm is expended, greater risk must be taken by utilizing germplasm resources that are more dissimilar. Here is one of the largest breeding concerns in applied breeding; dissimilar germplasm carries greater risk of dragging unwanted characters along with those that were desired, requiring more effort to develop the elite cultivars. Another related concern in acquiring the desired phenotype is whether the trait is qualitative or quantitative. Traits controlled by large numbers of genes need larger population sizes to locate a line that has all the required alleles. A monogenic, polymorphic trait has 2 combinations; a quantitative trait with 10 loci has 1024 combinations; and a quantitative trait with 20 loci has 1,048,756 combinations. Assuming 10 plants per meter², this is over 10 hectares if we also assume 1 and only 1 plant for every combination. So if one has, for arguments sake, 5 loci for fiber strength, 5 for length, 5 for micronaire, 2 for worm control, 1 for weed control, 10 for yield, 3 for plant habit, 4 for maturity, and 10 for disease control, this gives 2⁴⁴ or 17.6 trillion combinations. This does not exaggerate the task of the breeders; if anything it is understated. But, of course, a breeder doesn't do this all at once. Cotton breeders build the better combinations by combining the better elite types over the years and centuries to build even better cultivars. Any time that a breeder has to cross to poorly adapted types then every polymorphism with a less desired trait caused by the cross demands exponentially larger populations to get the desired phenotypic set back again. No wonder some breeders resist wide crosses.

In most respects, the general commercial breeding approaches for cotton have not changed much since the summary of Niles and Feaster (1984). The qualities that are desired in current cultivars are also not much different from those reported by Niles and Feaster, but there are differences in priorities. Yield is still the first three listed priorities (An old plant breeders joke and still true). Plant structure and maturity are key components within the broad topic of yield but several recent events in the cotton industry have changed how breeders look at plant size and earliness. The eradication of the boll weevil and the introduction of *Bt* cultivars have relaxed the need for earliness to combat the damage from boll weevils and worms. Therefore breeders are now revisiting the possibilities of utilizing a longer growing season so that the plant does not absolutely need an outstanding environment during the short reproductive period demanded by an early maturing crop. Early maturity and cold resistance during planting are still being pursued to expand the Cotton Belt into cooler regions with shorter growing seasons. The demand for a more compact growth habit has remained but much of the manner that is used to control vegetative growth has fallen on the use of Mepiquat-based plant growth regulators (PGR) (see

pubs.caes.uga.edu/caespubs/pubcd/B1305.htm for a general review). Cotton breeders are selecting against cotton lines that require excessive rates of Mepiquat-based PGRs to keep the crop from rank growth.

With the coming of herbicides, selectivity of control has been the desired goal. Even before herbicides, selectivity in weed control has been important; cultivator “blight” has been remarked upon for years. Until recently the burden of selectivity has been on the herbicide developers. With the advent of transgenic cotton with such traits as buctril-resistance, glyphosate-resistance, and glufosinate-resistance, non-selective herbicides have been able to be used for greater control of many somewhat related weed species, i.e. dicotyledonous weeds in dicotyledonous cotton. This has given producers a more labor efficient, economical system for weed control. As with any other herbicide weed control system, development of weeds with resistance to the herbicide is a realistic concern.

Fiber quality is getting more attention as the world cotton market is demanding better quality (Smith 2004) for use in modern spinning technology and to better compete with synthetic fibers. United States exports jumped from an average 7,033 bales (1994 to 2000) to 13,979 bales (2001 to 2007) (Anonymous 2007). However, over the last decade, the average quality of the U.S. Upland cotton has not improved, while quality requirements on export markets are constantly increasing (Estur 2004).

Breeding for drought tolerance and other environmental stresses is continuing to be important. One character that has gained in importance in the commercial breeding programs is environmental stability. High genotype by environment interaction directly leads to the need of specific cultivars for the different environmental areas. The expense of maintaining numerous cultivars with different transgenic trait packages and different seed treatment packages is not inconsequential. Commercial cotton breeders now strongly select for environmental stability across the Cotton Belt.

While the eradication of the boll weevil and the introduction of *Bt* cultivars have relaxed the need for earliness, it also has increased the importance of developing resistance to pests such as plant bugs and nematodes. Seed properties are also likely to receive a great deal more attention with the cost of crude oil and associated interest in biodiesel from plant oils. The variability of cottonseed oil in the National Cotton Variety Test ranges from 14% to 25% (Anonymous 2006) which shows that variability in oil content is available.

6.3 Developmental Breeding

In the beginning of Upland cotton production in the United States, the imported cultivars were not well adapted, but the variability in these plant populations was great enough that some survived (Ware 1951). This entire effort could be considered in the realm of developmental breeding to get around

the photoperiodic nature of these imports. Further natural crossing and selection produced cultivars that fit the annual production scheme of the temperate environments. Photoperiod sensitivity still is a great hindrance in utilizing wild cotton germplasm in breeding programs as well as in basic genetic research. Current research to develop these important germplasm sources as well as to understand the basic biology of photoperiodism would be facilitated by strong efforts to develop *Gossypium* as a model system directed towards understanding the genetic and physiological basis of photoperiodism.

Practical conversion of the photoperiodic stock into day-neutral breeding lines is ongoing. Day-neutrality is being introduced into numerous primitive *G. hirsutum* race stocks with a backcrossing program, USDA-ARS, Mississippi State University, Starkville, MS since the late 1970s (McCarty, Jenkins, Parrott, and Creech 1979; McCarty and Jenkins 1992; McCarty and Jenkins 1993; McCarty, Jenkins, and Robinson 1998; McCarty and Jenkins 2001; McCarty and Jenkins 2002; McCarty and Jenkins 2005b; McCarty and Jenkins 2005a; McCarty, Jenkins, and Wu 2005; McCarty, Wu, and Jenkins 2006). Procedurally, day-neutral F₂ offspring were selected from primitive race accessions that were crossed with a daylength-neutral parent and then backcrossed 4 times to the respective primitive accession. The converted race stocks have been found to be useful repositories of genetic diversity that are needed to enhance the genetic base of Upland cotton. This service of the USDA-ARS is of critical importance in maintaining the genetic advancement of cotton in the United States and throughout the world. Further research in this germplasm source is still needed so the desirable and novel agronomic traits can be placed into new commercial cultivars. Supplementary research has indicated that the converted lines from the primitive cottons indeed have favorable genes for yield and fiber quality (McCarty, Jenkins, and Zhu 1998a; McCarty, Jenkins, and Zhu 1998b; McCarty, Jenkins, and Wu 2004a; McCarty, Jenkins, and Wu 2004b; McCarty, Wu, and Jenkins 2007) and drought tolerance (Basal, Bebeli, Smith, and Thaxton 2003).

As the genetic distance further increases using the secondary and tertiary gene pools, further efforts are required and can be quite extreme working with the tertiary gene pool. In utilizing the secondary gene pool, *G. hirsutum* by *G. barbadense* hybrids are vigorous and fertile but have considerable levels of segregational breakdown (Beasley and Brown 1942; Stephens 1949; Hutchinson 1951). Even though direct use of hybrids between the allotetraploids is inefficient because of this hybrid breakdown as well as from partial sterility and likely later maturity of the offspring, rare successes were documented with the development of "Griffin" by John Griffin, Refuge Plantation, Greenville, MS and "Sealand" by W.H. Jenkins and associates, Florence, SC (Ware 1951). Both of these successes used systematic backcrosses; most of the other attempts that tried to directly combine *G. hirsutum* as Upland cotton with *G. barbadense* as Sea Island types showed that while the F₁ generation is attractive, the F₂ and F₃ generations were considered a "mess" and almost always discarded (Brown and Ware 1958). Further utilization of the Sea Island cotton via "Sealand"

has been seen in the use by the private breeding firms of the germplasm from the USDA-ARS Pee Dee program at Florence, SC (Bowman, May, and Calhoun 1996).

Upland cotton has a cytogenetic history of utilizing chromosomal anomalies such as monosomes (see Chapter 11 by Konan, Baudoin, D'Hont, and Mergeai, this volume) from which a series of alien chromosome substitutions lines may be synthesized (Endrizzi, Richmond, Kohel, and Brown 1963; White, Richmond, and Lewis 1967). These lines are useful to study the genetic effects of individual chromosomes on plant phenotypes as well as to help in estimating gene number, genetic interactions, and linkage relationships (Endrizzi, Turcotte, and Kohel 1985). With these substitutions, the exotic germplasm is manipulated on a chromosome basis instead of a genome-wide basis of direct combinations (Stelly, Saha, Raska, Jenkins, McCarty, and Gutierrez 2005; Jenkins, Wu, McCarty, Saha, Gutierrez, Hayes, and Stelly 2006; Saha, Jenkins, Wu, McCarty, Gutierrez, Percy, Cantrell, and Stelly 2006; Wu, Jenkins, McCarty, Saha, and Stelly 2006; Jenkins, McCarty, Wu, Saha, Gutierrez, Hayes, and Stelly 2007). The development of this type of series, however, requires special training, experience, and a lot of time; more than 20 years for a *G. hirsutum*/*G. barbadense* series (Wu et al. 2006) which is still not quite completed.

Although alien chromosome substitutions lines were one of the few ways to introgress other *Gossypium* tetraploids in a controlled manner into Upland cotton, the development of a scheme utilizing Advanced Backcross-QTL analysis (Tanksley and Nelson 1996) somewhat mitigates the difficulties found in direct interspecific crossing. By developing an extensive set of Near Isogenic Introgression lines from BC3 or BC4 families with each line characterized by one of a set of molecular markers that adequately spans the entire genome, one can evaluate relatively small segments of exotic germplasm for agronomic performance and analyze for QTLs. By targeting valuable genomic regions within an elite genetic background in this manner, fewer alleles need to be tested as well as having reduced linkage drag that can confound any analyses. Replicated testing of the initial 37 individual NILs developed from a *G. hirsutum* by *G. barbadense* (cv. Tamcot 2111 and cv. Pima S-6) show the efficacy of this approach by monitoring the fineness (mic range - 3.2 to 5.7), length (UHM range - 0.99 to 1.22 inches), strength (range - 26.4 to 33.3 g/tex), and % short fiber content (range - 6.7 to 10.8%) of the fiber (Lubbers, Chee, Paterson, and Smith 2006). Unfortunately, some utilization of the secondary gene pool is not possible since not all chromosomes appear to be retained as segregation occurs in this interspecific hybrid (Jiang et al. 2000). The Advanced Backcross-QTL schema is currently being exploited with additional *Gossypium* tetraploids and can be employed quickly while offering greater resolution in reducing the undesirable effects of linkage drag.

Further efforts are required to deploy the traits found in the diploid species of *Gossypium*. The most common effective procedures involve the development of either synthetic hexaploids or tetraploids that can then be manipulated into hybridizing with the AD genome of *G. hirsutum* in a recurrent backcrossing

scheme (Stewart 1995). These manipulations include the use of colchicine (Blakeslee and Avery 1937) to bridge from the diploids to the tetraploids via doubling chromosome sets to hexaploids and tetraploids that can somewhat freely hybridize with *G. hirsutum* and show recombination. Although Beasley (1940b; 1940a) wasn't the first to use poisons such as colchicine to double chromosome numbers in *Gossypium* (Mendes 1939 from Beasley 1940a), he is considered the principal contributor to improving the techniques of hybridizing and obtaining fertility in the offspring of interspecific crosses in cotton (Brown and Ware 1958).

In the tertiary gene pool, a fertile hybrid is not easily obtained and requires a higher level of manipulation to place the exotic, alien germplasm in a fertile system that can provide for recombination between the recipient and the donor chromosomes. The part of the diploids that include the A or D genome along with the B and F genomes (those genomes that are homoeologous to the A and D genomes) will probably recombine well once fertility is established. There are, however, several species in this group that carry complementary lethal genes (*G. davidsonii*, *G. klotzschianum*, *G. gossypoides*, and the sanguineum race of *G. arboreum*) (Stewart 1994; Mergeai 2003) but methodologies have been developed that keep these genes from being barriers to introgression (Lee 1981). The other part of the of the *Gossypium* diploids, genomes C, E, G, and K, are more difficult to obtain recombination as these diploids are considered to be more distantly related to the parental A and D genomes found in *G. hirsutum*. Strategies to improve the *G. hirsutum* germplasm via interspecific hybridization at Gembloux Agricultural University are described in a review by Mergeai (2003) and in his chapter of this volume. Stewart (1992) reviewed these techniques in enhancing disease resistance genetics.

6.4 Further Research Areas of Value

Information about changes in genetic diversity through time will be valuable in determining such things as the level of the bottleneck caused by strength of selection during the boll weevil invasion. Comparing the genetic diversity of the races would provide evidence regarding whether they indeed reflect incremental or branching stages of domestication. Molecular markers used in estimating diversity could also be used in QTL (quantitative trait loci) and linkage disequilibrium (LD) analyses. Although QTL analysis has been used since the mid 1990s to determine the genetics of different traits that are selected in breeding for cotton improvement, indispensable work in comparing germplasm groups such as 'yucatanense' with the other races is yet to be reported. Further racial categorization of the specimens in the cotton germplasm collection is also an important first step in evaluating the genetic distances within and between the races. Linkage disequilibrium analysis is also beginning to be used in cotton (Abdurakhmonov, Kohel, Saha, Pepper, Yu, Buriev, Abdullaev, Shermatov, Jenkins, Scheffler, and Abdurkarimov 2007; Myers and Badigannavar 2007) but needs to be directly applied to the domestication process of Upland cotton.

7 Summary

The diversity of the *G. hirsutum* germplasm base is currently narrow. However there are many sources of diversity available from the primary, secondary, and tertiary gene pools. The 7 races of *G. hirsutum* are directly usable as a germplasm resource with photoperiodism as the main barrier. The sister tetraploids require a little more effort to be fully utilized due to segregational breakdown. The diploids generally require more extreme methods to be introgressed such as chromosome doubling and the use of bridging species. The phenotypic consequences of the domestication of Upland cotton can be described by characters found in the domestication syndrome that is common to many crop plants. These are also the leading traits to study in order to illuminate the genetics of domestication of Upland cotton.

Upland cotton has been a leading crop in research on utilizing distantly related germplasm and the required tools are available, but much more effort is needed. Applied breeding to bring novel elements of both qualitative and quantitative traits from the germplasm base into the commercial arena is necessary for the cotton industry to survive and thrive. Basic work on the genetic tools necessary to efficiently manipulate the germplasm is also imperative for continuing progress in the future.

References

- Abdurakhmonov, I.Y., Kohel, R.J., Saha, S., Pepper, A.E., Yu, J.Z., Buriev, Z.T., Abdullaev, A., Shermatov, S., Jenkins, J.N., Scheffler, B., and Abdurakimov, A. (2007). Genome-wide linkage disequilibrium revealed by microsatellite markers and association study of fiber quality traits in cotton. In: *Plant & Animal Genomes XV Conference* (San Diego, CA, USA: Scherago).
- Anonymous. (2006). 2006 National Cotton Variety Test (Stoneville, MS: USDA-ARS, Crop Genetics & Production Research Unit, <http://www.ars.usda.gov/SP2UserFiles/Place/64021500/2006NCVT.pdf>), pp. 139.
- Anonymous. (2007). Cotton - Summary Data - United States (National Cotton Council of America, <http://www.cotton.org/econ/cropinfo/cropdata/summary.cfm>).
- AOGTR. (2002). The biology and ecology of cotton (*Gossypium hirsutum*) in Australia (Australian Office of the Gene Technology Regulator, <http://www.agbios.com/docroot/decdocs/06-059-003.pdf>), pp. 30.
- Applequist, W.L., Cronn, R., and Wendel, J.F. (2001). Comparative development of fiber in wild and cultivated cotton. *Evolution & Development* 3, 3–17.
- Basal, H., Bebeli, P., Smith, C.W., and Thaxton, P. (2003). Root growth parameters of converted race stocks of upland cotton and two BC₂F₂ populations. *Crop Science* 43, 1983–1988.
- Beasley, J.O. (1940a). The production of polyploids in *Gossypium*. *Journal of Heredity* 31, 39–48.
- Beasley, J.O. (1940b). Hybridization of American 26-chromosome and Asiatic 13-chromosome species of *Gossypium*. *Journal of Agricultural Research* 60, 175–181.
- Beasley, J.O., and Brown, M.S. (1942). Asynaptic *Gossypium* plants and their polyploids. *Journal of Agricultural Research* 1, 421–427.

- Blakeslee, A.F., and Avery, A.G. (1937). Methods of inducing doubling of chromosomes in plants. *Journal of Heredity* 28, 393–411.
- Bowman, D., May, O.L., and Calhoun, D.S. (1997). Coefficients of parentage for 260 cotton cultivars released between 1970 and 1990 (Washington, DC: USDA-ARS, Tech. Bull. no. 1852).
- Bowman, D.T., May, O.L., and Calhoun, D.S. (1996). Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Science* 36, 577–581.
- Bridge, R.R., Meredith, W.R., Jr., and Chism, J.F. (1971). Comparative performance of obsolete varieties and current varieties of Upland cotton. *Crop Science* 11, 29–32.
- Briggs, F.N., and Knowles, P.F. (1967). *Introduction to plant breeding*. (New York: Reinhold Publishing Corporation).
- Brown, H.B., and Ware, J.O. (1958). *Cotton*. (New York: McGraw-Hill).
- Brubaker, C.L., and Wendel, J.F. (1994). Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear Restriction-Fragment-Length-Polymorphisms (RFLPs). *American Journal of Botany* 81, 1309–1326.
- Calhoun, D.S., Bowman, D.T., and May, O.L. (1994). Pedigrees of Upland and Pima cotton cultivars released between 1970 and 1990. *Mississippi Agricultural and Forestry Experiment Station Bulletin* 1017.
- Clement, C.R. (1999). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany* 53, 188–202.
- Culp, T.W., and Harrell, D.C. (1975). Influence of lint percentage, boll size, and seed size on lint yield of Upland cotton with high fiber strength. *Crop Science* 15, 741–746.
- Damp, J.E., and Pearsall, D.M. (1994). Early cotton from coastal Ecuador. *Economic Botany* 48, 163–165.
- Dillehay, T.D., Rossen, J., Andres, T.C., and Williams, D.E. (2007). Preceramic adoption of peanut, squash, and cotton in northern Peru. *Science* 316, 1890–1893.
- Doebley, J. (2004). The genetics of maize evolution. *Annual Review of Genetics* 38, 37–59.
- Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321.
- Duggar, J.F. (1907). Descriptions and classification of varieties of American Upland cotton (Auburn, AL: Agricultural Experiment Station of The Alabama Polytechnic Institute Bull. 140), pp. 24.
- Endrizzi, J.E., Turcotte, E.C., and Kohel, R.J. (1985). Genetics, cytology, and evolution of *Gossypium*. In: *Advances in Genetics* (Academic Press, Inc.), pp. 271–375.
- Endrizzi, J.E., Richmond, T.R., Kohel, R.J., and Brown, M.S. (1963). Monosomes-A tool for developing better cottons. *Tex. Agric. Prog.* 9, 9–11.
- Estur, G. (2004). Quality requirements on export markets for U.S. cotton (Washington, DC: International Cotton Advisory Committee, http://www.icac.org/cotton_info/speeches/estur/2004/quality_reqs_us_exp.pdf).
- Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J.P., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science* 289, 85–88.
- Freire, E.C., Moreira, J.A.N., Santos, J.W., and Andrade, F.P. (1998). Relações taxonômicas entre os algodoeiros Moco e *Gossypium mustelinum* do nordeste brasileiro. *Pesqui. Agropecu. Bras.* 33, 1555–1561.
- Fryxell, P.A. (1965). A revision of the Australian species of *Gossypium* with observations on the occurrence of *Thespesia* in Australia (Malvaceae). *Australian Journal of Botany* 13, 71–102.
- Fryxell, P.A. (1979). *The natural history of the cotton tribe*. (College Station, TX: Texas A&M University Press).
- Fuller, D.Q. (2007). Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the old world. *Annals of Botany* 100, 903–924.

- Gulati, A.M., and Turner, A.J. (1928). A note on the early history of cotton. Indian Central Cotton Committee Technical Laboratory Bulletin 17.
- Harlan, J.R. (1992). *Crops and man*. (Madison, WI: American Society of Agronomy-Crop Science Society of America).
- Harlan, J.R., and de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20, 509–517.
- Harris, D.R. (1989). An evolutionary continuum of people-plant interaction. In: *Foraging and Farming, The evolution of plant exploitation* (London: Unwin Hyman Ltd.), pp. 11–26.
- Hutchinson, J.B. (1951). Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* 5, 161–193.
- ICAC. (2004). Cotton: Review of the world situation (Washington D. C.: International Cotton Advisory Committee), pp. 20.
- Iqbal, M.J., Pepper, A.E., El-Zik, K.M., and Reddy, O.U.K. (2001). A genetic bottleneck in the 'evolution under domestication' of Upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics* 103, 547–554.
- Jenkins, J.N., Wu, J.X., McCarty, J.C., Saha, S., Gutierrez, O., Hayes, R., and Stelly, D.M. (2006). Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in topcrosses with Upland cotton cultivars: I. Yield and yield components. *Crop Science* 46, 1169–1178.
- Jenkins, J.N., McCarty, J.C., Wu, J., Saha, S., Gutierrez, O., Hayes, R., and Stelly, D.M. (2007). Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in topcrosses with Upland cotton cultivars: II. Fiber quality traits. *Crop Science* 47, 561–570.
- Jiang, C., DelMonte, T.A., Paterson, A.H., Wright, R.J., and Woo, S.S. (2000). QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theoretical and Applied Genetics* 100, 409–418.
- Kislev, M.E., Hartmann, A., and Bar-Yosef, O. (2006). Early domesticated fig in the Jordan Valley. *Science* 312, 1372–1374.
- Knight, R.L., and Hutchinson, J.B. (1950). The evolution of blackarm resistance in cotton. *J. Genet.* 50, 36–58.
- Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396.
- Lacape, J.M., Dessauw, D., Rajab, M., Noyer, J.L., and Hau, B. (2007). Microsatellite diversity in tetraploid *Gossypium* germplasm: Assembling a highly informative genotyping set of cotton SSRs. *Molecular Breeding* 19, 45–58.
- Lee, J.A. (1981). Genetics of D₃ complementary lethality in *Gossypium hirsutum* and *Gossypium barbadense*. *Journal of Heredity* 72, 299–300.
- Lee, J.A. (1984). Cotton as a world crop. In: *Cotton*, R.J. Kohel and C.F. Lewis, (Eds.) (Madison, WI: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.), pp. 1–25.
- Lewton, F.L. (1925). Notes on the taxonomy of American and Mexican Upland cottons. *J. Washington Academy of Sciences* 15, 65–71.
- Li, C.B., Zhou, A.L., and Sang, T. (2006). Rice domestication by reducing shattering. *Science* 311, 1936–1939.
- Lubbers, E.L., Chee, P.W., Paterson, A.H., and Smith, C.W. (2006). Fiber quality of a near-isogenic introgression line series from an Upland by Pima interspecific cross. In: *Beltwide Cotton Conferences* (San Antonio, TX), pp. 845–852.
- Lubbers, E.L., Chee, P.W., May, O.L., Gannaway, J.R., and Paterson, A.H. (2005). Genetic relationships of historically important eastern U.S. Upland cotton. In: *Beltwide Cotton Conferences* (New Orleans, LA), pp. 1027–1030.
- Main, T.F. (1912). Cambodia cotton in Bombay Presidency. *Agricultural Journal of India* 7, 373–381.

- May, O.L., Bowman, D.T., and Calhoun, D.S. (1995). Genetic Diversity of Us Upland Cotton Cultivars Released between 1980 and 1990. *Crop Science* 35, 1570–1574.
- McCarty, J.C., and Jenkins, J.N. (2005a). Registration of 14 primitive derived cotton germplasm lines with improved fiber strength. *Crop Science* 45, 2668–2669.
- McCarty, J.C., and Jenkins, J.N. (2005b). Registration of 21 day length-neutral flowering primitive cotton germplasm lines. *Crop Science* 45, 2134–2134.
- McCarty, J.C., Jenkins, J.N., and Robinson, M. (1998). Cotton germplasm: Root-knot nematode resistance in day-neutral primitive accessions (Mississippi State University: Mississippi Agricultural & Forestry Experiment Station Research Report vol 22 no 3), pp. 4.
- McCarty, J.C., Jenkins, J.N., and Zhu, J. (1998a). Introgression of day-neutral genes in primitive cotton accessions: I. Genetic variances and correlations. *Crop Science* 38, 1425–1428.
- McCarty, J.C., Jenkins, J.N., and Wu, J.X. (2004a). Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: II. Genetic effects and genotypic values. *Crop Science* 44, 1231–1235.
- McCarty, J.C., Jenkins, J.N., and Wu, J.X. (2004b). Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. *Crop Science* 44, 1226–1230.
- McCarty, J.C., Jenkins, J.N., and Wu, J. (2005). Potential of primitive accessions for cotton improvement (Mississippi State University: Office of Agricultural Communications, Division of Agriculture, Forestry, and Veterinary Medicine, Mississippi State University; MAFES Bull. 1141), pp. 22.
- McCarty, J.C., Wu, J., and Jenkins, J.N. (2006). Genetic diversity for agronomic and fiber traits in day-neutral accessions derived from primitive cotton germplasm. *Euphytica* 148, 283–293.
- McCarty, J.C., Wu, J., and Jenkins, J.N. (2007). Use of primitive derived cotton accessions for agronomic and fiber traits improvement: Variance components and genetic effects. *Crop Science* 47, 100–110.
- McCarty, J.C., Jenkins, J.N., Parrott, W.L., and Creech, R.G. (1979). The conversion of photoperiodic primitive race stocks of cotton to day-neutral stocks (Mississippi State University: Mississippi Agricultural and Forestry Experiment Station Research Report vol 4 no 19), pp. 4.
- McCarty, J.C., Jr., and Jenkins, J.N. (1992). Cotton germplasm: Characteristics of 79 day-neutral primitive race accessions (Mississippi State University: Mississippi Agricultural and Forestry Experiment Station, Tech Bull no 184), pp. 17.
- McCarty, J.C., Jr., and Jenkins, J.N. (1993). Registration of 79 day-neutral primitive cotton germplasm lines. *Crop Science* 33, 351–351.
- McCarty, J.C., Jr., and Jenkins, J.N. (2001). Primitive cotton germplasm: Yield and fiber traits for 16 day-neutral accessions (Mississippi State University: Mississippi Agricultural & Forestry Experiment Station, Research Report vol 22 no 19), pp. 5.
- McCarty, J.C., Jr., and Jenkins, J.N. (2002). Registration of 16 day length-neutral flowering primitive cotton germplasm lines. *Crop Science* 42, 1755–1756.
- McCarty, J.C., Jr., Jenkins, J.N., and Zhu, J. (1998b). Introgression of day-neutral genes in primitive cotton accessions. II. Predicted genetic effects. *Crop Science* 38, 1428–1431.
- Mendes, A.J.T. (1939). Duplicacao do numero de cromosomios em cafe, algodao e fumo, pela acao de colchicina. In: *Boletim Tecnico No. 57* (Sao Paulo, Brazil: Secretaria da Agricultura, Industria, e Comercio), pp. 21.
- Merzeai, G. (2003). Forty years of genetic improvement of cotton through interspecific hybridisation at Gembloux Agricultural University: Achievement and prospects. In: *Proc. World Cotton Research Conference - 3. Cotton Production for the New Millennium*, A. Swanepoel, ed (Cape Town, South Africa: Agricultural Research Council, Institute for Industrial Crops, Pretoria), pp. 120–133.

- Miller, P.A., and Rawlings, J.O. (1967). Selection for increased lint yield and correlated responses in Upland cotton, *Gossypium hirsutum* L. *Crop Science* 7, 637–640.
- Myers, G.O., and Badigannavar, A. (2007). Association genetics for quantitative characters in tetraploid cotton (*G. hirsutum* L.). In: *2007 Beltwide Cotton Conferences* (New Orleans, Louisiana: National Cotton Council of America).
- NCCoA. (2006). World cotton database (National Cotton Council of America, <http://www.cotton.org/econ/cropinfo/cropdata/index.cfm>).
- Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution* 29, 1–10.
- Niles, G.A., and Feaster, C.V. (1984). Breeding. In: *Cotton*, R.J. Kohel and C.F. Lewis, (Eds.) (Madison, WI: ASA/CSSA/SSSA), pp. 202–229.
- Paterson, A.H. (2002). What has QTL mapping taught us about plant domestication? *New Phytologist* 154, 591–608.
- Pickersgill, B. (1977). Taxonomy and origin and evolution of cultivated plants in New World. *Nature* 268, 591–595.
- Pickersgill, B., Barrett, S.C.H., and Andrade-Lima, D.d. (1975). Wild cotton in northeast Brazil. *Biotropica* 7, 42–54.
- Pillay, M., and Myers, G.O. (1999). Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Science* 39, 1881–1886.
- Rieger, R., Michaelis, A., and Green, M.M. (1976). *Glossary of genetics and cytogenetics: classical and molecular*. (Berlin: Springer-Verlag).
- Saha, S., Jenkins, J.N., Wu, J.X., McCarty, J.C., Gutierrez, O.A., Percy, R.G., Cantrell, R.G., and Stelly, D.M. (2006). Effects of chromosome-specific introgression in Upland cotton on fiber and agronomic traits. *Genetics* 172, 1927–1938.
- Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z.C., Tai, Y.S., Gill, B.S., and Faris, J.D. (2006). Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172, 547–555.
- Smith, R. (2004). To improving fiber quality: Cotton breeding goals have shifted (New York, NY: Southwest Farm Press, Penton Media, Inc., http://www.southwestfarmpress.com/mag/farming_improving_fiber_quality/).
- Stelly, D.M., Saha, S., Raska, D.A., Jenkins, J.N., McCarty, J.C., Jr., and Gutierrez, O.A. (2005). Registration of 17 Upland (*Gossypium hirsutum*) Cotton Germplasm Lines Disomic for Different *G. barbadense* Chromosome or Arm Substitutions. *Crop Sci* 45, 2663–2665.
- Stephens, S.G. (1949). The cytogenetics of speciation in *Gossypium*. I. Selective elimination of the donor parent genotype in interspecific backcrosses. *Genetics* 34, 627–637.
- Stephens, S.G. (1958). Factors affecting seed dispersal in *Gossypium* and their possible evolutionary significance (North Carolina Agricultural Experiment Station Tech. Bull. No. 131), pp. 32.
- Stephens, S.G. (1965). The effects of domestication on certain seed and fiber properties of perennial forms of cotton, *Gossypium hirsutum* L. *The American Naturalist* 99, 355–372.
- Stephens, S.G. (1967). Evolution under domestication of the New World cottons (*Gossypium* spp). *Ciencia E Cultura* 19, 118–134.
- Stephens, S.G. (1972). Geographical distribution of cultivated cottons relative to probable centers of domestication in the New World. In: *12th International Latin American Symposium*, A.M. Srb, ed (Cali, Colombia: Plenum Press), pp. 39–254.
- Stephens, S.G. (1976). Some observations on photoperiodism and the development of annual forms of domesticated cottons. *Economic Botany*, 409–418.
- Stephens, S.G., and Phillips, L.L. (1972). The history and geographical distribution of a polymorphic system in New World cottons. *Biotropica* 4, 49–60.
- Stewart, J.M. (1992). Germplasm resources and enhancement strategies for disease resistance. In: *Beltwide Cotton Conferences*, D.J. Heber and D.A. Richter, eds (Memphis, TN: National Cotton Council of America), pp. 1323–1325.

- Stewart, J.M. (1994). Potential for crop improvement with exotic germplasm and genetic engineering. In: *Challenging the Future: Proceedings of the World Cotton Research Conference-1*, G.A. Constable and N.W. Forrester, eds (Brisbane, Australia: CSIRO, Melbourne, Australia), pp. 313–327.
- Stewart, J.M., and Kerr, T. (1974). Relationship between fiber-length increase and seed-volume increase in cotton (*Gossypium hirsutum* L.). *Crop Science* 14, 118–120.
- Tanksley, S.D., and Nelson, J.C. (1996). Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics* 92, 191–203.
- Tatineni, V., Davis, D.D., and Cantrell, R.G. (1996). Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Science* 36, 186–192.
- Townsend, T. (2006). Personal communication (Washington DC: International Cotton Advisory Committee).
- Tyler, F.J. (1910). Varieties of American Upland cotton (USDA-ARS, BPI, Bulletin no. 163).
- UNCTAD. (2006). Cotton: Characteristics (United Nations Conference on Trade and Development).
- Van Becelaere, G., Lubbers, E.L., Paterson, A.H., and Chee, P.W. (2005). Pedigree- vs. DNA marker-based genetic similarity estimates in cotton. *Crop Science* 45, 2281–2287.
- Van Esbroeck, G.A., Bowman, D.T., Calhoun, D.S., and May, O.L. (1998). Changes in the genetic diversity of cotton in the USA from 1970 to 1995. *Crop Science* 38, 33–37.
- Van Esbroeck, G.A., Bowman, D.T., May, O.L., and Calhoun, D.S. (1999). Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. *Crop Science* 39, 323–328.
- Vaughan, D.A., Balazs, E., and Heslop-Harrison, J.S. (2007). From crop domestication to super-domestication. *Ann Bot* 100, 893–901.
- Wang, H., Nussbaum-Wagler, T., Li, B.L., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. *Nature* 436, 714–719.
- Ware, J.O. (1951). *Origin, rise, and development of American Upland cotton varieties and their status at present*. (Fayetteville, AR: University of Arkansas, College of Agriculture, Agricultural Experiment Station).
- Wendel, J.F., Brubaker, C.L., and Percival, A.E. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *American Journal of Botany* 79, 1291–1310.
- White, T.G., Richmond, T.R., and Lewis, C.F. (1967). Use of cotton monosomes in developing interspecific substitution lines (Washington DC: Crop Res. USDA Rep. ARS 34–91), pp. 15.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics* 16, 97–159.
- Wu, J., Jenkins, J., McCarty, J., Saha, S., and Stelly, D. (2006). An additive-dominance model to determine chromosomal effects in chromosome substitution lines and other gemplasms. *Theoretical and Applied Genetics* 112, 391–399.
- Zohary, D., and Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World. *Science* 187, 319–327.
- Zohary, D., and Hopf, M. (2000). *Domestication of plants in the old world: The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley*. (New York: Oxford University Press).

The Worldwide Gene Pool of *Gossypium barbadense* L. and Its Improvement

Richard G. Percy

Abstract This chapter describes the improved and unimproved gene pools of *Gossypium barbadense*. Section one discusses the taxonomic and geographic structure of species diversity. Section two describes the origin and development of modern improved germplasm pools, beginning with Sea Island cottons developed in the Caribbean and the coastal Southeast of the United States in the late 18th century. The origins and development of the Egyptian and Amercian Pima germplasm pools are sketched. Finally, smaller but significant gene pools created in Peru, Israel, and Australia are discussed. The role of interspecific introgression with *G. hirsutum* in broadening the improved germplasm pool is noted. In section three unimproved genetic resources *in situ* and in germplasm collections are examined.

1 The Structure of Species Diversity

Morphological and ecological variants of *G. barbadense* have led to various binomials being applied to the species over the years. At various times at least 15 synonyms have been applied to portions of the species (Fryxell 1979). In older literature the most common synonyms are *G. vitifolium*, *G. peruvianum*, *G. brasiliense*, *G. acuminatum* and *G. darwinii*. These putative species were often discriminated on the basis of a small suite of morphological traits that were often simply inherited and tended to intergrade. With the exception of *G. darwinii*, all these species names have been reduced to the single *G. barbadense* species. The lack of distinctive, discontinuous variation within the species is perhaps demonstrated by the fact that old species names have not proven useful enough to be retained as varietal names. Of the many species names, only *brasiliense* and *darwinii* have been retained as varieties by some (Hutchinson, Silow, and

R.G. Percy (✉)

USDA-ARS, Southern Plains Agricultural Research Center, College Station, Texas, USA
e-mail: Richard.Percy@ars.usda.gov

Stephens 1947; Fryxell 1979), whereas *G. darwinii* has been reinstated as a species by others (Wendel and Percy 1990).

Greater success has been realized in ascertaining geographical structure in *G. barbadense* species' diversity. Observations by Mauer (1930) and Harland (1936) placed the center of diversity of the species west of the Andean cordillera of Columbia, while Hutchinson et al. placed the center of variability in the mountain valleys of northwestern South America (Hutchinson, Silow, and Stephens 1947). An investigation of species diversity using isozyme analysis revealed the greatest genetic diversity in the species (and probable center of origin) in northwestern South America, west of the Andes (Percy and Wendel 1990). Analyses of regional diversity revealed that the West of Andes region had a higher genetic identity with the East of Andes region than it did with either the Caribbean or Central American regions. The regional genetic diversity of the species progressively declined from the West of Andes region through the East of Andes, Caribbean, and Central American regions. Taken together, this suggests a dispersal route of *G. barbadense* out of its West of Andes center of origin through eastern South America and the Caribbean to Central America.

Investigators have placed the center of domestication of *G. barbadense* in coastal Peru and Ecuador, based upon archeological evidence (Rossen, Dillehay, and Ugent 1996; Piperno and Pearsall 1998) and observed gradients in plant morphological traits consistent with trends toward domestication (Stephens and Moseley 1974). A recent survey of genetic diversity within primitive South American germplasm using AFLP fingerprinting has refined this center of early domestication to coastal regions of northwestern Peru and southwestern Ecuador around the Gulf of Guayaquil (Westengen, Huaman, and Heun 2005). Allozyme evidence suggest that modern cultivated *G. barbadense* derives from germplasm acquired directly from the centers of origin, diversity, and domestication of the species, while historical records report some contributions being made from germplasm originating east of the Andes (Watt 1907).

2 Origins and Development of Modern Improved Germplasm

2.1 Sea Island Cottons

Although the origins of modern cultivated *Gossypium barbadense* are complex and somewhat obscure, it appears that most modern gene pools derive directly or indirectly from Sea Island cottons – many by way of Egyptian germplasm. Sea Island cottons, originally found in the Caribbean Islands, were introduced to mainland North America in the late 18th century (Fryxell 1965) and became the basis of a small but flourishing industry in the coastal states of the Southeastern United States (McGowan 1961). From their early beginnings, Sea Island cottons were distinguished by their exceptionally long and fine fiber – qualities that both

placed a premium price on their fiber and limited market demand. The Sea Island industry in the U.S. persisted through the 19th century into the 20th century. However, by the 1920's it was in serious decline. Prior to this ultimate decline, cottons from South Carolina, Georgia, and Florida were used to re-establish a Sea Island industry in the West Indies. Today, Sea Island cottons residing in germplasm collections are often identified by their island of origin (St. Vincent, Barbados, St. Lucia, St. Kitts, Nevis, etc.) or by their fiber traits (St. Vincent Superfine, OSI Ordinary, Superfine V46, etc.).

Although first mentioned in historical records as growing in the Caribbean (Fryxell 1965), allozyme evidence places Sea Island cottons in closest affinity with cottons of coastal Peru and Ecuador, west of the Andes (Percy and Wendel 1990). This affinity suggests that Sea Island cotton arose from an introduced source, rather than from *G. barbadense* native to the Caribbean. Fryxell states that the photoperiod neutral flowering habit found in today's commercial cottons ultimately derives from Sea Island cottons (Fryxell, 1979). It is not known when Sea Island cottons acquired the photoperiod neutral flowering habit. However, there is no historical record to indicate that the Sea Island cottons introduced to North America were not already photoperiod neutral. It is reasonable to postulate that the acquisition and selection for photoperiod neutrality was accompanied by a severe restriction in genetic variability. Countering this postulated constriction in diversity is evidence of variability introduced through introgression with *G. hirsutum* (Percy and Wendel 1990; Wang, Dong, and Paterson 1995). In fact, several investigators have postulated that an introgression event was responsible for the creation of the Sea Island germplasm (Watt 1907; Kerr 1960; Stephens 1975; Stephens 1976). Stephens (1975) postulated that photoperiod neutrality itself was acquired from *G. hirsutum* in an introgressive event. The presence of conserved, site specific *G. hirsutum* chromatin in Sea Island cottons and other improved *G. barbadense* germplasm (Wang, Dong, and Paterson 1995) has led to speculation that these introgressed segments play a role in conferring beneficial traits – such as fiber quality or photoperiod neutrality.

Improvement efforts in Sea Island cottons began in the eighteenth century and continued into the twentieth century. The first recorded selection and improvement efforts were for earlier maturity and annual growth habit, which were necessitated by the introduction of Sea Island cottons to the North American mainland (Hutchinson 1959). Not surprisingly, however, much of the breeding in Sea Island cottons has been for the improvement or maintenance of fiber quality. Hutchinson has described a system practiced within St. Vincent V 135 in the early twentieth century in which 20 individual plants were selected annually for their fiber type, their seed planted in progeny rows, and selected progeny rows bulked for two years to produce planting seed. Seed from production fields was discarded and crushed for oil. This system, while not breeding per se, was highly successful in maintaining the extreme uniformity of fiber length and fineness necessary for the commercial production of Sea Island cotton. On the mainland United States, wilt caused by *Fusarium oxysporum* Schlecht. F. sp. *Vasinfestum* became a severe problem at the

beginning of the twentieth century. Selection of resistant individuals within heavily infested fields proved successful in identifying and isolating highly resistant sources (Ware 1936). Two *Fusarium* resistant cultivars resulting from these efforts were Rivers and Centerville. Working in the West Indies, the Empire Cotton Growing Corporation began to apply modern breeding principles and techniques to Sea Island cottons in 1943 (Hutchinson and Manning 1943). As a group, Sea Island cottons have retained their exceptional fiber length and fineness, but are exceedingly late maturing and low in productivity in comparison with their Pima and Egyptian antecedents.

2.2 Egyptian Cottons

The creation of a modern cotton production and marketing system in Egypt is usually dated to 1820. In that year the Khedive Mohammed Ali, at the instigation of a French engineer named M. Jumel, fostered the commercial production of a cotton variety known as “Jumel” (Watt 1907; Balls 1919). Jumel was perennial tree cotton, producing brown lint, which was imported to Egypt from Sudan. There is confusion concerning the origins of Jumel’s tree cotton. Watt (1907) refers to Jumel as being *Gossypium brasiliense*. Formerly given species status, *brasiliense* is a form of *G. barbadense* found in eastern South America. Balls (1919) identifies Jumel as being a Peruvian-type cotton originating from west of the Andes. Dunn (1949) stated that Jumel was “a variety of the herbaceum... an Asiatic type”, which is undoubtedly wrong. Regardless of its origins, the production of Jumel was successful enough to stimulate the importation and experimentation with Sea Island cultivars and cultivars from Brasil. It appears that these introduced cottons interbred among themselves and with Jumel in an uncontrolled manner to produce a population from which the first Egyptian cultivar Ashmouni was derived in 1860 (Kearney 1943). Ashmouni was described as being a very heterogeneous cultivar that produced brown lint. Ashmouni in turn gave rise to the cultivar Mit Affifi. Mit Affifi was derived from a single plant selection made in a field of Ashmouni in 1882. The parent plant may have been an “off-type” resulting from a natural cross of Ashmouni with Sea Island (Dunn 1949). Many subsequent cultivars have been derived from Mit Affifi through selection or from natural crosses of Mit Affifi with Sea Island and other cultivars (Kearney 1943). Cultivars resulting from selection or “discovery” in fields of Mit Affifi include Nubari (1905), Assili (1906), and Sakel (1906). Other early Egyptian cultivars supposedly resulting from natural hybridization include: Abbasi, which was supposedly derived from a natural cross of Mit Affifi and Zafiri; Giza 3, which was supposedly derived from an Ashmouni x Sakel hybrid population; and Giza 7, which was supposedly derived from the same natural hybrid cross that produced Giza 3 (Kearney 1943; Dunn 1949). By the 1930’s the Egyptian Ministry of Agriculture was creating new cultivars through controlled hybridization and selection. Examples of early

cultivars developed through deliberate, controlled hybridization include: Karnak, derived from a cross of Maarad and Sakha 3, released in 1931; Menoufi, a selection from the cross between Giza 12 and Sakha 3, released in 1942; and Giza 30, a product of the cross of Giza 7 with Sakha 11, a Sea Island derivative.

Throughout the early development of Egyptian cottons, selection and improvement concentrated upon productivity, fiber quality and regional adaptation. Although several distinct environments were recognized in Egypt, these environments are essentially subsets of three broader regions – Lower Egypt or the Delta, Middle Egypt, and Upper Egypt. Kearney in 1943 stated that the hotter climate of Upper Egypt required cotton grown in that region to have “the ability to withstand high temperatures without excessive boll-shedding”. Therefore, heat tolerance has been a significant factor among other environmental adaptations bred for. From the inception of the Egyptian cotton industry, its cotton has been marketed to a high fiber quality, specialty market. Selection for fiber quality in Egyptian cottons has evolved over time into the development of two general fiber types – long staple (LS) and extra long staple (ELS). By the late 1980’s most of the Egyptian extra long staple cultivars were being grown in the cooler Delta region close to the coast. Long staple cultivars were being grown in more inland regions of the Delta and in Middle and Upper Egypt. The regional division of Egyptian cultivars by fiber types reflects historical efforts by the Egyptian government to control marketing and prevent surpluses of cotton of a particular staple. However, this targeting of genotypes to specific regions also may be a reflection of the fact that the extra long staple Sea Island cottons, which played a significant role in the development of Egyptian ELS cultivars, were more heat sensitive and were poorly adapted to many inland Egyptian environments.

Another concern in early improvement efforts of Egyptian cotton was the development of Fusarium wilt resistant cultivars. The disease was first observed around 1903, but it was not until 1923 that active selection for resistance was begun. Ironically, the progenitor of Egyptian cottons, Ashmouni, was reported to be immune to the disease (Dunn 1949). However, this resistance had been lost in many cultivars – particularly those having the susceptible cultivar Sakel as a parent. Sakha 4 was the first success at producing a long staple resistant cultivar. Giza 3, Giza 7, and Wafeer, all reported to have been derived from Sakel X Ashmouni hybrids, were found to be resistant.

2.3 American Pima Cottons

Prior to 1900, three separate attempts were made to introduce Egyptian cotton into the United States (Ware 1936). The first attempt in 1867 was the direct result of the disruption of trade created by the American Civil War – which had shifted much of the international cotton market to Egyptian cottons. The superior fiber quality of Egyptian cottons had found considerable favor with mills, and it was believed that to recapture the market the U. S. would have to

grow cotton of comparable fiber quality. The attempt in 1867 failed, as did subsequent attempts in 1892 and 1897. In the failed 1897 introduction attempt, testing of Egyptian cottons was extended to the irrigated river valleys of the southwestern U.S. where growing conditions were considered to be similar to the Nile Valley. As a result of these tests, another more successful attempt at introduction was made in Arizona and California between 1900 and 1902 using the cultivars Mit Afifi, Ashmouni, Abbasi, and Jannovich, among others (McGowan, 1961). Early breeding efforts focused on “acclimatizing” the Egyptian germplasm through individual plant selection within these cultivars. In 1907 individual plant selections were made within Mit Afifi that would lead to the release in 1912 of the first American-Egyptian cultivar, named Yuma (Smith, Cantrell, Moser, and Oakley, 1999). Yuma was earlier, more productive, and had greater fiber length than its Mit Afifi parent (McGowan 1961) and competed favorably with Egyptian imports for most uses (Scofield, Kearney, Brand, Cook, and Swingle 1919). In 1910, during the development of Yuma, a plant selection was made within it that would become the cultivar named Pima. Released in 1918, Pima had a longer and finer fiber than that of Yuma (Smith et al 1999) or any Egyptian cultivar of the time (Brown 1953). However its fiber, along with its parent’s fiber, was reported to be weaker than those of the better Egyptian cultivars of the time. This did not deter Egyptian breeders from using Pima, renamed Maarad, as a parent in the development of the cultivar Karnak (Kearney 1943). Pima was to be the principal ELS cultivar grown in the U.S. for the following 15 years. In 1918, at the time of the release of Pima, a cross was made between it and the Egyptian variety Sakel (Kearney 1943). The resulting cultivar, SxP, slowly replaced Pima as the dominant ELS cultivar of the American Southwest between 1935 and 1941. Another cultivar, Amsak, was the product of crossing SxP back to Sakel in an attempt to improve fiber quality. Due to lower yield potential, Amsak was not widely accepted or grown. A final American-Egyptian cultivar, Pima 32, was released in 1948 (Peebles 1950). Pima 32 was the result of crossing an SxP x Pima progeny with the Egyptian cultivar Giza 7.

In summary, three Egyptian cultivars, Mit Afifi, Sakel, and Giza 7, were the progenitors of the American-Egyptian germplasm pool, commonly referred to retrospectively as the Sacaton gene pool (Fig. 1). Although considerable progress was made in adapting Egyptian germplasm to the southwestern U.S., the gene pool was fairly narrow, due to its derivation from a narrow base (Kerr 1960, Smith et al. 1999). This situation changed with the creation of the Pima S-1 and Hybrid-B gene pools.

The Pima S-1 gene pool was created between 1934 and 1947 by the intercrossing of three diverse *G. barbadense* germplasm sources and introgression from *G. hirsutum* (Peebles, 1954; Feaster and Turcotte, 1962; Smith et al. 1999). The three *G. barbadense* sources included the American-Egyptian Pima, a Tanguis parent of Peruvian origin, and a Sea Island parent. The fourth parent was an unnamed *G. hirsutum* cultivar from the Stoneville germplasm (Fig. 1). The series of crosses that resulted in the creation of the Pima S-1 cultivar are

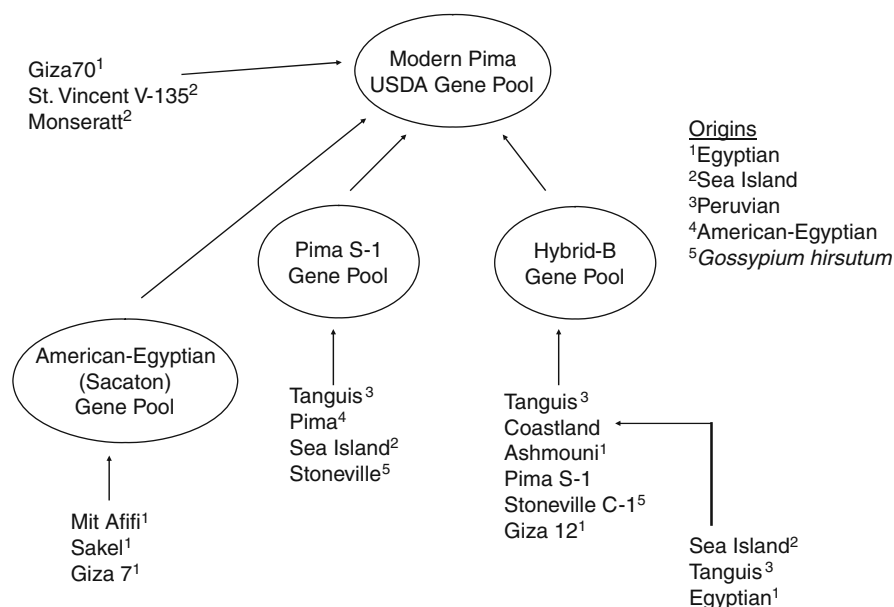


Fig. 1 Sources of the modern USDA Pima gene pool

uncertain. However, the resulting original Pima S-1 was a very heterogeneous mixture of genotypes. From this mixture, 160 strains were selected and evaluated. Of the 160 strains, nine were bulked to create the finished Pima S-1 cultivar. Pima S-1, its component strains, and its derivatives became the Pima S-1 gene pool. Whereas the Pima S-1 gene pool was created and developed in hot, low desert environments of Arizona, another gene pool was being created in a higher elevation, more moderate environment at El Paso Texas (Feaster, Turcotte, and Young 1967). Like the Pima S-1 gene pool, the Hybrid-B gene pool resulted from the interbreeding of diverse *G. barbadense* germplasm sources and introgression from *G. hirsutum*. These sources included Peruvian Tanguis, Egyptian Ashmouni and Giza 12, Pima S-1, Stoneville C-1, and Coastland. The latter contributor was itself the product of crossing among Sea Island, Tanguis, and Egyptian sources. Only one cultivar, Pima S-3, was derived from the Hybrid-B gene pool (Young, Feaster, and Turcotte 1976). Genotypes from the Hybrid-B gene pool, developed under more moderate environments at higher elevations, generally exhibited sensitivity to higher temperatures present in low desert environments.

The modern USDA Pima gene pool resulted from the combining of the American-Egyptian or Sacaton gene pool, the Pima S-1 gene pool, and the Hybrid-B gene pool (Fig. 1). The first cultivar to derive from this combined gene pool was Pima S-2 whose parents Pima S-1 and 3-79 derived from the Pima S-1 and Sacaton gene pools, respectively. Successive cultivars displayed

varying contributions from the three gene pools, but the Pima S-1 gene pool predominated. Pima S-3 was an exception, deriving solely from the Hybrid-B gene pool. Pima S-4 and Pima S-5 derived from crosses among lines of the Pima S-1 and American-Egyptian gene pools (Feaster and Turcotte 1976). The cultivar Pima S-6, released in 1983, derived from all three gene pools. In the pedigree of Pima S-6, the PS-1 gene pool is represented five times, the American-Egyptian gene pool three times, and the Hybrid-B gene pool once (Feaster and Turcotte 1984). Although the USDA gene pool originated from diverse sources and was considered to have a genetically broad base, the pool was essentially closed from the mid 1950s until the late 1980s. Beginning in 1989, Sea Island and Egyptian germplasm were re-introduced into the USDA gene pool – resulting in the release of higher fiber quality germplasm lines in 1998 and 2002 (Percy 1998; Percy 2002).

As in any cotton breeding program, increased productivity and fiber quality have been primary objectives of improvement efforts in American Pima. The achievement of higher productivity under the high temperature environments common in much of the American Southwest has led to an emphasis on selection for heat tolerance or heat escape. Early in the development of the American-Egyptian breeding efforts emphasis was placed on “adapting” Egyptian cultivars to the southwest. Adaptation, to a great degree, corresponded with the acquisition of heat tolerance or heat avoidance. In the modern USDA germplasm, traits associated with early maturity, increased determinacy, and therefore heat avoidance, such as height of first fruiting branch, plant height at season’s end, and flowering rate have been emphasized in the development of Pima cultivars (Feaster and Turcotte 1965; Feaster, Young, and Turcotte 1980). Although traits associated with early maturity and determinacy were found to be advantageous under heat stress inducing environments, these same traits were observed to be neutral or even disadvantageous in the absence of heat stress (Feaster and Turcotte 1965). Selection for productivity under heat stress conditions in the Pima germplasm for over fifty years has led to the inadvertent selection for increased levels of stomatal conductance in successive Pima cultivars (Radin, Lu, Percy, and Zeiger 1994; Lu, Percy, Qualset, and Zeiger 1998). Apparently, higher stomatal conductance levels have led to higher evaporative cooling rates and heat avoidance under irrigated conditions in the arid Southwest. With the shift in recent years of Pima production away from low desert localities to the more moderate Mediterranean environment of California, Pima cultivar development has tended toward slightly later maturing, more indeterminate genotypes.

The shift in Pima production to California in the past decade has resulted in another breeding objective arising. Beginning around 2003, severe losses to *Fusarium* wilt were observed in isolated fields in that state. Pima cultivars were found to be particularly susceptible, and this susceptibility was found to be associated with the presence of the previously unidentified race 4 of the pathogen (Kim, Hutmacher, and Davis 2005; Hutmacher, Davis, Ulloa, Wright, Munk, Vargas, Roberts, Marsh, Keeley, Kim, and Percy 2005). Efforts to identify sources of resistance within Pima have met with some success and are

continuing. Genetic investigations of resistant Pima genotypes indicate that resistance is simply inherited and may be conditioned by one or two genes (Ulloa, Hutmacher, Davis, Wright, Percy, and Marsh 2006). Present breeding efforts include broadening the genetic base of Fusarium resistant germplasm.

2.4 Other Improved Germplasm Pools

A modern extra-long staple cotton industry in Peru appears to have originated with the introduction of the Egyptian Mit Afifi cultivar to that country in the late 19th or early 20th centuries (McGowan 1961). In 1923 the American-Egyptian cultivar Pima was introduced in Peru and by 1933 it had replaced the acclimatized Mit Afifi. The Egyptian cultivars Sakel and Karnak were also introduced and grown in Peru in the twentieth century (Brown 1938), where they were presumably selected within and acclimatized. In the 1990's the Peruvian extra-long staple industry was again introducing American Pima germplasm. At that time an active program of hybridization of American Pima and native Tanguis cultivars was initiated, resulting in the release of a new cultivar in 2006 (Lazo 2006). Throughout the history of extra-long staple cotton production in Peru, cottons of superior fiber length have been produced. Whether this is due to re-selection within introduced germplasm, advantageous environmental conditions, or a combination of the two is not clear.

Tanguis cottons, a second and distinct germplasm pool, have existed in Peru since 1918. Although the origins of Tanguis are uncertain, it is known that it was "discovered" on a plantation in 1915 and may be the result of a natural hybridization of native Peruvian cotton with an imported cultivar (Brown 1938; McGowan 1961). Molecular analyses of cultivated cottons support the contention that Tanguis cottons have a native origin in Peru, clustering most closely with an unimproved *G. barbadense* "check" accession in an "ancestral" cluster (Wang et al. 1995). Tanguis cottons possess distinctive fiber characteristics and are characterized by having extremely coarse fiber with high dye ability. Another distinction among Tanguis cottons is the presence of resistance to native races of Fusarium wilt that has been spread widely among cultivars through breeding and selection efforts. As previously mentioned, Tanguis cottons have contributed to the development of the modern American Pima germplasm.

An Israeli *G. barbadense* germplasm pool is the direct result of efforts to find resistance to Fusarium wilt. In the 1970's there was an extra-long staple cotton production industry in Israel that centered around the growing of American Pima. By the late 1970's this industry was in danger of collapsing due to severe losses to *Fusarium oxysporum* f.s. *vastinfectum* race 3. A resistance breeding program was established in 1978 utilizing seed from surviving plants of Pima S-4 and Pima S-5 from heavily infested fields (Fishler 2007). Highly resistant plants were identified among progeny of Pima S-4. These plants, possessing white pollen, inferior fiber properties, and excessive plant height, did not resemble Pima S-4, but rather an interspecific hybrid or perhaps an Egyptian

ancestor of Pima. Due to the agronomically inferior nature of these progeny, no effort was made to produce a cultivar from them through direct selection. Resistant progeny were crossed to Pima S-5 and after several generations of selection a resistant cultivar, F-27, was developed. Subsequent intercrossings of resistant progeny and Pima S-5 have resulted in the development and release of the cultivars BF-19 and PF-15. In the late 1990's the Israeli breeding program expanded its germplasm base by introducing more recent American Pima germplasm, Egyptian Giza 75, and materials of unknown origin.

Another, small, but unique *G. barbadense* germplasm pool has come into existence as the result of efforts to grow extra-long staple cotton in Australia. A major obstacle to production in Australia has been the presence of bacterial blight (*Xanthomonas campestris* pv *malvacearum* (Smith) Dye). This obstacle led Australian breeders to initiate a breeding program to transfer bacterial blight resistance to ELS cottons. Resistance gene transfer was accomplished in several ways. In one transfer, begun in 1992, the American Pima cultivar S-7 was crossed to the bacterial blight resistant *G. hirsutum* cultivar, Sicala V-1 (Constable 2007). Progeny of this cross were backcrossed approximately eight times (with screening for resistance in each backcross) to several American Pima lines to produce four separate families. In a second transfer, the resistant line Yseleta B2B6 was crossed to Pima S-7 and screened for bacterial blight resistance during three generations of selection to produce the line 95201. The parentage of Yseleta B2B6 is not known, but its appearance suggests that it is a product of *G. hirsutum* introgression. Another adversity that ELS cottons have faced in Australia is a particularly virulent race of *Fusarium* that occurs in certain areas. Screening among American Pima germplasm has identified levels of resistance to this race in some lines, including the line 8810. Interbreeding among bacterial blight and *Fusarium* resistant Australian lines, American Pima germplasm, and other introductions continues in the Australian germplasm.

2.5 Current Breeding Methods and Goals

Recurring themes in the breeding of *G. barbadense* are selection for fiber quality, heat tolerance, and *Fusarium* resistance. A constant and historical goal has been the increase of yield potential while maintaining fiber quality. Today, the vast majority of breeding efforts in *G. barbadense* employ pedigree breeding, with hybridization followed by several generations of selection within and among progeny. In some programs early generation selection is omitted and selection is first conducted in the F₃ or F₄ generation populations composed of more homozygous individuals. Another common trend in the breeding history of *G. barbadense* has been purposeful introgression with *G. hirsutum*. Introgression has been followed by direct selection or by the initiation of backcrossing when a specific trait of interest has been transferred. There have been attempts to produce intraspecific F₁ hybrids in *G. barbadense*, but the

resulting hybrids have generally not produced the heterotic response seen in interspecific hybrids and therefore have been less economically viable (Davis 1978). Greater heterotic responses and therefore greater economic potential have been observed in interspecific hybrids of *G. barbadense* and *G. hirsutum*. Viable interspecific hybrid breeding programs are currently found in India, China, and Israel. A trend within *G. barbadense* breeding programs in recent years has been increased efforts to broaden germplasm pools by introducing parental materials from multiple sources outside the immediate pool. As seen in the above discussion, available improved *G. barbadense* germplasm pools, while often divergent in fiber traits, growth habits, stress tolerances, etc., share common ancestry. While the common ancestry of improved germplasm pools would seem to imply a genetic narrowness, results from molecular investigations suggest otherwise.

2.6 Diversity within Cultivated *G. barbadense* and the Role of Introgression

In an allozyme investigation of the diversity within *G. barbadense*, Percy and Wendel (1990) found that a group of 26 improved lines (including American Pima, Egyptian, Russian, and Peruvian Tanguis cultivars), possessed genetic variation equivalent to that found in unimproved accessions collected from the center of diversity west of the Andes in Ecuador and Peru. Percy and Wendel attributed much of the diversity within improved cultivars of *G. barbadense* to introgression with *G. hirsutum*, and reported that 22% of the cultivars sampled had *G. hirsutum* alleles. Another investigation, using restriction fragment length polymorphisms, found *G. hirsutum* chromatin in all of 54 *G. barbadense* cultivars tested (Wang et al. 1995). This latter report stated that *G. hirsutum* chromatin regions within *G. barbadense* cultivars did not appear to be random but that there were specific sites that appeared to be conserved. As previously mentioned, Sea Island cottons, the progenitors of almost all modern cultivars, are thought by some to be the product of introgression. The report of Wang et al. (1995) found that the frequency of *G. hirsutum* alleles were highest in Sea Island cultivars and lowest in Peruvian cultivars. Regardless of the diversity within *G. barbadense* cultivars due to *G. hirsutum* introgression, the large variation occurring in unimproved germplasm remains a reservoir to be exploited.

3 In Situ and Ex Situ Germplasm Resources

A major portion of the intraspecific variation in *G. barbadense* remains in unimproved germplasm and is an under-utilized resource. Several reasons exist for this situation, including the short-day photoperiodism, perennial growth habit, and poor agronomic and fiber traits of much of the unimproved

germplasm. These limitations are not insurmountable obstacles in accessing the diversity present in the species, however. Of graver concern is the loss of *in situ* species diversity due to human activities. Although the concept of *in situ* conservation of genetic resources has gained recognition in recent years, its practice presents unique challenges. Habitat loss due to conversion of wild lands to agricultural use is probably the greatest challenge and is most problematic in developing nations (Rubenstein, Heisey, Shoemaker, Sullivan, and Frisvold 2005). Although habitat loss affects whole plant and animal communities as well as individual species, it is often easier to justify and gain support for community preservation than it is to justify conservation of a single wild species of little apparent economic value. Another source of genetic diversity and therefore a candidate for *in situ* conservation are landraces (also, in the case of cotton, dooryard or commensal cottons). Since landrace or dooryard cottons frequently are grown in subsistence or basic economies, the adoption of a modern cotton industry has often led to their abandonment in favor of improved cultivars (Singh, Mohan, Kulkarni, Baitule, and Pathak 2003). Unlike wild species that are sometimes the beneficiaries of benign neglect in a wild environment, there is little chance of landrace preservation *in situ* on agricultural land without some form of economic incentive being offered (Rubenstein et al., 2005). Wild species also can be adversely affected by the adoption of modern cotton production practices in adjacent territories. Wild and feral populations of cotton often harbor the same insect and disease pests as adjoining cultivated cotton, are perceived as pest sources, and become the target of eradication programs. In such situations, *in situ* management of wild species may become necessary, as well as some forms of compensation and economic incentive for *in situ* preservation to be successful. Few examples of the implementation of active *in situ* preservation efforts currently exist for cotton. One example is an effort to protect native colonies of the species *G. mustelinum* in Brazil (Barroso, P.A., Batisita, L.V., Hoffman, L.V. and Ciampi, A.Y. 2006). This one conservation effort demonstrates that, with local support, conservation can be as simple as erecting a fence to exclude cattle.

Given the challenges faced by *in situ* conservation, it is fortunate that *ex situ* resources exist in the form of national germplasm collections. According to the U.N. FAO "Report on the State of the World's Plant Genetic Resources" of 1996 (U.N. FAO 1996b), over 49,000 accessions of all species of *Gossypium* were then residing in institutional *ex situ* germplasm collections. Of 41 institutions that reported having cotton related activities and participated in the World Information and Early Warning System of the FAO, 25 reported having *ex situ* germplasm collections. Currently, the largest collections are maintained in the USA, China, Russia, India, France, and Brasil (Table 1). Together, these six collections contain around 33,000 accessions – or approximately two thirds of all accessions held *ex situ*. Data on the composition of collections is readily available for the USDA-ARS collection of the United States (USDA-ARS,NGRP,GRIN 2007) and the CIRAD collection of France (Dessauw and Hau, 2007). As one might expect, the species *G. hirsutum* has the highest

Table 1 Size of the six largest *ex situ* cotton germplasm collections of the world

Country	Germplasm Accessions (all species)
	no.
India	4,500 [‡]
France	3,070 [‡]
Russian Federation	6,300 [§]
United States	9,303 [¶]
China	6,724 [#]
Brazil	2,835 [‡]
total	32,732

[‡] Dessau and Han, 2007[§] U.N. FAO, 1996a[¶] USDA-ARS, NGRP, GRIN, 2007[#] Weidong et al., 2000

representation in these two collections – comprising 57% of the USDA collection and 71% of the CIRAD collection. *G. barbadense* accessions comprise a much smaller portion of the two collections, or about 16–17 %. These fractions of *G. hirsutum* and *G. barbadense* are probably typical of most national germplasm collections. Currently the USDA-ARS collection contains around 1,626 accessions of *G. barbadense* originating from 64 countries. Of the total number of accessions, 1,447 can be identified by country of origin. Approximately 660 of the accessions originate from countries within the native, ancestral range of *G. barbadense*. The majority of these 660 accessions are commensal or wild cottons. Of these accessions, 359 originate from Peru and Ecuador, where the centers of origin and diversity of the species reside. One hundred and fifty-eight of the accessions within the collection can be identified as cultivars, obsolete cultivars, or land races. A further 303 of the accessions can be identified as being improved breeding lines. The improved germplasm pools that have contributed significantly to modern ELS cottons are represented in the U.S. National Cotton Germplasm Collection by 63 Sea Island accessions, 34 Egyptian accessions, 271 American Pima accessions, and 17 Peruvian Tanguis accessions. The smaller French CIRAD collection contains 483 *G. barbadense* accessions. Of these 483 accessions, 214 are cultivars and 214 are wild or commensal cottons. The improved germplasm pools that have contributed significantly to modern ELS cottons are represented in the French CIRAD Collection by 41 Sea Island accessions, 55 Egyptian accessions, 16 American Pima accessions, and 34 Peruvian Tanguis accessions. The Peruvian Aspero and African Mono cottons, which should be regarded as landraces, are represented in the CIRAD collection by 18 and 35 accessions, respectively.

In conclusion, a large part of the genetic variability observed in modern *G. barbadense* cultivars may be due to introgression with *G. hirsutum*. Recent efforts to expand the variability of improved germplasm pools have emphasized interbreeding among elite improved *G. barbadense* gene pools and introgression

with *G. hirsutum*. Significant genetic variability exists in the landrace, wild or commensal germplasm pools held in collections or existing *in situ*. These resources remain under-utilized, although there have been recent discovery efforts within them. The status of *in situ* genetic resources in many regions is becoming increasingly precarious. Cooperative public and private conservation and exchange of germplasm is essential to maintaining a healthy cotton improvement industry.

References

- Balls, W.L. (1919) *The Cotton Plant in Egypt*. MacMillan and Co., London.
- Barroso, P.A., Batisita, C.E., Hoffman, L.V. and Ciampi, A.Y. (2006) Genetic structure and *in situ* conservation of natural populations of *Gossypium mustelinum*. Int. Cotton Genome Initiative Res. Conf. Brasilia, Brazil.
- Brown, C.H. (1953) *Egyptian Cotton*. Leonard Hill Ltd., London.
- Brown, H.B. (1938) *Cotton*. McGraw-Hill, New York.
- Constable, G. (2007) Program Leader, Cotton Management and Improvement, CSIRO, Australia, personal communication.
- Davis, D.D. (1978) Hybrid cotton – specific problems and potentials. Adv. Agron. 30, 129–157.
- Dessaui, D. and Hau, B. (2007) Inventory and history of the CIRAD cotton (*Gossypium* spp.) germplasm collection. Plant Gen. Res. Newsletter 147, 53–58. <http://www.ipgri.cgiar.org/pgnew/sletter/archive.asp>
- Dunn, R.P. (1949) *Cotton in Egypt*. National Cotton Council, Memphis, TN.
- Feaster, C.V., and Turcotte, E.L. (1962) Genetic basis for varietal improvement of Pima cottons. USDA-ARS Bull. 34–31.
- Feaster, C.V., and Turcotte, E.L. (1965) Fruiting height response: a consideration in varietal improvement of Pima cotton, *Gossypium barbadense* L. Crop Sci. 5, 460–464.
- Feaster, C.V., and Turcotte, E.L. (1976) Registration of Pima S-4 cotton. Crop Sci. 16, 604.
- Feaster, C.V., and Turcotte, E.L. (1976) Registration of Pima S-5 cotton. Crop Sci. 16, 604.
- Feaster, C.V., and Turcotte, E.L. (1984) Registration of Pima S-6 cotton. Crop Sci. 24, 382.
- Feaster, C.V., Turcotte, E.L. and Young, E.F. (1967) Pima cotton varieties for low and high elevations. USDA-ARS Bull. 34–90.
- Feaster, C.V., Young, E.F. and Turcotte, E.L. (1980) Comparison of artificial and natural selection in American Pima cotton under different environments. Crop Sci. 20, 555–559.
- Fishler, G. 2007. The history and performance of the Eden Fusarium resistant Pima varieties (compiled from Hebrew reports 1986–1987). Personal communication.
- Fryxell, P.A. (1965) Stages in the evolution of *Gossypium* L. Advan. Front. Pl. Sci. 10, 31–56.
- Fryxell, P.A. (1979) *The Natural History of the Cotton Tribe*. Texas A&M University Press, College Station and London.
- Harland, S.C. (1936) Some notes on cotton in Columbia. Trop. Agric. Trinidad 13, 1–13.
- Hutchinson, J.B. (1959) *The Application of Genetics to Cotton Improvement*. Cambridge University Press, Cambridge.
- Hutchinson, J. B. and Manning, H.L. (1943) The efficiency of progeny row breeding in cotton improvement. Emp. J. Exp. Agr. 11, 140.
- Hutchinson, J.B., Silow, R.A., and Stephens, S.G. (1947) *The Evolution of Gossypium*. Oxford University Press, London, New York, Toronto.
- Hutmacher, R.B., Davis, R.M., Ulloa, M., Wright, S., Munk, D.S., Vargas, R.N., Roberts, B.A., Marsh, B.H., Keeley, M.P., Kim, Y. and Percy, R.G. (2005) Fusarium in Acala and Pima cotton: symptoms and disease development. Proc. Beltwide Cotton Conf., New Orleans, LA, pp. 245–246.

- Kearney, T.H. (1943) Egyptian-type cottons: their origin and characteristics. Report of Division of Cotton and Other Fiber Crops and Diseases, USDA Mimeo (unnumbered).
- Kerr, T. (1960) The potentials of barbadense cottons. Proc. 12th Ann. Cotton Imp. Conf., Memphis, TN, pp 57–60.
- Kim, Y., Hutmacher, R.B. and Davis, R.M. (2005) Characterization of California isolates of *Fusarium oxysporum* f. sp. *vasinfectum*. Plant Dis. 89, 366–372.
- Lazo, J. (2007) Ficha Técnica de la variedad Pima-IPA-59. Release notice of Instituto Peruano del Algodón – IPA (Peruvian Cotton Institute).
- Lu, Z., Percy, R.G., and E. Zeiger. (1998) Stomatal conductance predicts yields in irrigated Pima cotton and bread wheat grown at high temperatures. J. Exp. Bot. 49, 453–460.
- Mauer, F.M. (1930) The cottons of Mexico, Guatemala, and Colombia. Bull. Appl. Bot. Genet. Pl. Breed. Suppl. 47, 543–553.
- McGowan, J.C. (1961) *History of Extra-long Staple Cottons*. Hill Printing Co., El Paso, Tex.
- Peebles, R.H. (1950) 1950 long staple picture changed. Cotton Trade J., April 14, 7–8.
- Peebles, R.H. (1954) Current status of American-Egyptian cotton breeding. Proc. 6th Cotton Imp. Conf. 1954, 1–8.
- Percy, R.G. (1998) Registration of extra-long staple cotton germplasm, 89590 and 8810. Crop Sci. 38, 1407.
- Percy, R.G. (2002) Registration of five extra-long staple cotton germplasm lines possessing superior fiber length and strength. Crop Sci. 42, 988.
- Percy, R.G. and Wendel, J.F. (1990) Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor. Appl. Genet. 79, 529–542.
- Piperno, D.R. and Pearsall, D.M. (1998) *The Origins of Agriculture in the Lowland Neotropics*. Academic Press, San Diego.
- Radin, J.W., Lu, Z., Percy, R.G., and Zeiger, E. (1994) Genetic variation for stomatal conductance in Pima cotton and its relation to improvements of heat adaptation. Proc. Natl. Acad. Sci. 91, 7217–7221.
- Rossen, J., Dillehay, T.D., and Ugent, D. (1996) Ancient cultigens or modern intrusions? Evaluating plant remains in an Andean case study. J. Archaeol. Sci. 23, 391–407.
- Rubenstein, K.D., Heisey, P., Shoemaker, R., Sullivan, J. and Frisvold, G. (2005) Crop genetic resources. An economic appraisal. USDA-ERS, Economic Information Bull. 2. <http://www.ers.usda.gov/Publications/EIB2/>
- Scofield, C.S., Kearney, T.H., Brand, C.J., Cook, O.F., and Swingle, W.T. (1919) Production of American Egyptian cotton. USDA Bull. 742.
- Singh, V.V., Mohan, P., Kulkarni, V.N., Baitule, S.J. and Pathak, B.R. (2003) Explorations within India for collection of cotton species germplasm. Plant Gen. Res. Newsletter 136, 40–46. <http://www.ipgri.cgiar.org/pgrnewsletter/archive.asp>
- Smith, C.W., Cantrell, R.G., Moser, H.S., and Oakley, S.R. (1999) History of cultivar development in the United States. In: C.W. Smith and J.T. Cothren (Eds.), *Cotton*. John Wiley, New York, pp. 99–171.
- Stephens, S.G. (1975) Some observations on photoperiodism and the development of annual forms of domesticated cottons. Econ. Bot. 30, 409–418.
- Stephens, S.G. (1976) The origin of Sea Island cotton. Agric. Hist. 50, 391–399.
- Stephens, S.G. and Moseley, M.E. (1974) Early domesticated cottons from archaeological sites in central coastal Peru. Am. Antiquity 39, 109–122.
- Ulloa, M., Hutmacher, R.B., Davis, R.M., Wright, S.D., Percy, R.G., and Marsh, B. (2006) Breeding for Fusarium Wilt race 4 resistance in cotton under field and greenhouse conditions. J. Cotton Sci. 10, 114–127.
- U.N. FAO (1996b) Report on the state of the world's plant genetic resources. FAO, Rome, Italy. <http://www.fao.org/ag/AGP/AGPS/Pgrfa/pdf/SWRFULL2.PDF>
- USDA-ARS-National Genetic Resources Program (2007) Germplasm Resources Information Network (GRIN) database. <http://www.ars-grin.gov/cgi-bin/npgs/html/taxfam.pl>

- Wang, G.L., Dong, J.M. and Paterson, A.H. (1995) The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germ plasm: molecular analysis of introgressive plant breeding. *Theor. Appl. Genet.* 91, 1153–1161.
- Ware, J.O. (1936) Plant breeding and the cotton industry. In: *Yearbook of Agriculture*. United States Dept. Agriculture, Government Printing Office, Washington, DC, pp. 657–744.
- Watt, G. (1907) *The Wild and Cultivated Cotton Plants of the World*. Longmans, Green, and Co. London, New York, Bombay, and Calcutta.
- Wendel, J.F. and Percy, R.G. (1990) Allozyme diversity and introgression in the Galapagos-Islands endemic *Gossypium darwinii* and its relationship to continental *Gossypium barbadense*. *Biochem. Syst. Ecol.* 18, 517–528.
- Weidong, G., Fang, J., Zheng, D., Li, Y., Lu, X., Rao, R., Hodgkin, T. and Zongwen, Z. (2000) Utilization of germplasm conserved in Chinese national genebanks – a survey. *Plant Gen. Res. Newsletter* 123, 1–8 <http://www.ipgri.cgiar.org/pgrnewsletter/archive.asp>
- Westengen, O.T., Huaman, Z., and Heun, M. (2005) Genetic diversity and geographic pattern in early South American cotton domestication. *Theor. Appl. Genet.* 110, 392–402.
- Young, E.F., Feaster, C.V., and Turcotte, E.L. (1976) Registration of Pima S-3 cotton. *Crop Sci.* 16, 604.

The Worldwide Gene Pools of *Gossypium arboreum* L. and *G. herbaceum* L., and Their Improvement

Venkatesh N. Kulkarni, Basavaraj M. Khadi, Manjula S. Maralappanavar, Lalitadas A. Deshapande, and S. S. Narayanan

Abstract Among the four cultivated species of *Gossypium*, diploid species (*G. herbaceum* and *G. arboreum*) are generally cultivated in marginal, drought prone environments of Asia due to their inherent ability to withstand drought. Diploid cottons are also known for their ability to resist sucking pests such as hoppers, white flies, thrips and aphids, and leaf curl virus. Before the spread of *G. hirsutum* in Asia, diploids were grown widely, but gradually replaced due to short, coarse and weak fiber and undesirable boll and plant features. These are still cultivated in traditional cotton growing areas of India, Pakistan, China, Bangladesh and Iran. India has structured genetic improvement programs aiming to improve yield potential and fiber quality of diploids. World germplasm collections are preserved in India, however, small sets of local/imported collections are also maintained in France, USA, China, and Iran. *G. herbaceum* has contributed the A-genome to tetraploid (AADD) cotton and is important to studying evolution of cultivated cotton and fiber related genes. The origin, evolution and distribution of A-genome diploids have been presented in this chapter along with different local types available in India. Compiled information on these important gene pools of cotton is seldom found in the literature. We made efforts to assemble valuable information on breeding efforts made in developing new genotypes or biotechnologically important diploid cotton populations.

1 Introduction

Gossypium arboreum L. and *G. herbaceum* L. are the two diploid ($2n=2x=26$) cultivated cotton species popularly known as Old World, Asiatic or Desi cotton (in India and Pakistan). We refer to these as diploid cultivated cotton, which

V.N. Kulkarni (✉)

Senior Cotton Breeder, JK Agri-Genetics Ltd, 1-10-177, 4th Floor, Varun Towers, Begumpet, Hyderabad, India – 500 016
e-mail: vnkulkarni@jkseeds.net

belong to the A-genome (Fryxell 1979, 1992). *Gossypium herbaceum* has contributed the cytoplasm and nuclear A-genome to tetraploid ($2n=4x=52$) cotton (Wendel 1989; Chapter 1 of this volume by Wendel, Brubaker, Alvarez, Cronn, and Stewart), which are referred to as New World cotton. Morphologically, tetraploid and diploid cultivated cotton differ in plant habit, leaves, flower, boll, seed and fiber characters. Although the diploid cultivated cotton species have overlapping morphological characters, both can be distinguished to some extent based on plant habit, leaves, bracteole and boll features (Hutchinson, Silow and Stephens 1947; Fryxell 1979; 1992). They are annual subshrubs in habit with *G. arboreum* taller and lankier than *G. herbaceum*. Long and narrow lobed leaves are common in *G. arboreum*, whereas a slightly constricted lobe base is normal in *G. herbaceum* leaves making it a broad lobed type. Bracteoles with fewer (3–4) teeth are closely invested to buds that are longer than broad in *G. arboreum*, while *G. herbaceum* has more teeth (6–8) flaring widely from the buds. In *G. arboreum*, bolls are mostly round with tapering points and a less smooth to deeply pitted rough boll surface. Round bolls with less/no tapering beak and smooth to less pitted surfaces are features of *G. herbaceum*. Individually, such characters may not help in unambiguous distinction of the two species, however, taken as a group reliable distinctions can often be made.

Genetic, cytogenetic, isozyme and molecular investigations, all clearly suggest that *G. herbaceum* and *G. arboreum* are two distinct species. These two species can be easily crossed to produce F_1 's that are fertile and vigorous with pollen fertility of about 60%. Due to genetic incompatibility, in the F_2 and subsequent generations, higher frequencies of non-viable seeds, seedling mortality and abnormal plant morphologies are common. Good plants appearing in the segregating generations tend to resemble one of the parents. Such observations suggest that these species differ distinctly. Cytogenetically, these two species differ by one reciprocal translocation (Gerstel 1953; Gerstel and Sarvella 1956; Phillips 1961). Isozyme variation at 40 loci among 103 *G. arboreum* and 31 *G. herbaceum* accessions provide additional evidence of species distinctness as there were eight polymorphic loci unique to *G. herbaceum* and 13 unique to *G. arboreum* (Wendel, Olson, and Stewart 1989). Molecular evidence for species distinctness comes from diversity analysis using AFLP markers in cultivated Indian diploid cotton, where *G. herbaceum* and *G. arboreum* formed two different clusters (Rana and Bhat 2004). In cluster analysis based on RAPD markers the genotypes of the respective species also grouped into distinct clusters, providing additional support of species distinctness of *G. arboreum* and *G. herbaceum* (Rana and Bhat 2002).

Cotton breeders desire to use these diploid genetic resources to enhance the biotic and abiotic stress tolerance of cultivated tetraploid cotton. In the Indian sub-continent, efforts are in progress to genetically improve the cultivated diploids for plant, boll and fiber features. Several diploid cultivated cotton accessions/improved lines are invaluable genetic resources for biotechnological studies. Accordingly, we have made an effort to elucidate global diploid cultivated gene pools with regard to their evolution, domestication, geographical

distribution, available germplasm resources, relationship with other species, and breeding implications. We also discuss the future uses of these cotton species.

2 Origin and Evolution of Diploid Cultivated Cotton

Among the 14 Afro-Arabian-Asiatic species belonging to four genomes (A, B, E and F), *G. herbaceum* and *G. arboreum* were domesticated for human use. All analyses conducted to date suggest that the four groups represent four independent evolutionary lineages. Seelanan, Schnabel and Wendel (1997), based on a “molecular clock”, estimated that the A-genome diverged from the other three-genome groups approximately 4 to 9 million years ago. Within the A-genome lineage, *G. herbaceum* and *G. arboreum* separated more recently. Domesticated *G. herbaceum* was derived from its wild ancestor *G. herbaceum* var *africanum* (Watt) Hutch and Ghose. It has been suggested that *G. arboreum* was derived from domesticated *G. herbaceum* based on (a) study of archeological remains attributable to *G. arboreum* discovered at the Mohenjo-Daro site (Gulati and Turner 1928); (b) cytologically, the genomes of *G. arboreum* and *G. herbaceum* are distinguished by a reciprocal translocation (Gerstel 1953) suggesting that *G. arboreum* arose as an incipient species with the origin and fixation of the translocation; and (c) morphological similarity existing between *G. arboreum* race ‘indicum’ (a primitive perennial form in western India, ‘Rozi’ cotton) and *G. herbaceum* race ‘acerifolium’ and ‘wightianum’. Based on the magnitude of differences in genetic diversity between *G. herbaceum* and *G. arboreum* (measured by isozyme analysis), Wendel, Olson and Stewart (1989) concluded that the two species must have diverged from each other circa 1–4 million years ago, i.e. before human evolution, implying that the two species were domesticated independently from divergent wild progenitors. Hence, evidence regarding the origin and evolution of *G. arboreum* remains somewhat obscure.

3 Domestication and Distribution of Diploid Cultivated Cotton

3.1 *Gossypium herbaceum*

Cultivated *G. herbaceum* derived from a truly wild diploid ancestor, *G. herbaceum* var *africanum* (Hutchinson 1959), which occurs naturally in Southern Africa. It was assumed that traders introduced this wild form into Southern Arabia, where the first domestication of Old World cotton took place. From here bi-directional spread occurred, one west into Africa and other east into western India (Fig. 1). The northward spread of *G. herbaceum* var *africanum* might have followed loss of photoperiodism. The primitive cultivated forms of *G. herbaceum* are the perennial race ‘acerifolium’ that arose in Southern Arabia and the Iranian coastal area, and then were further distributed by Arab traders in

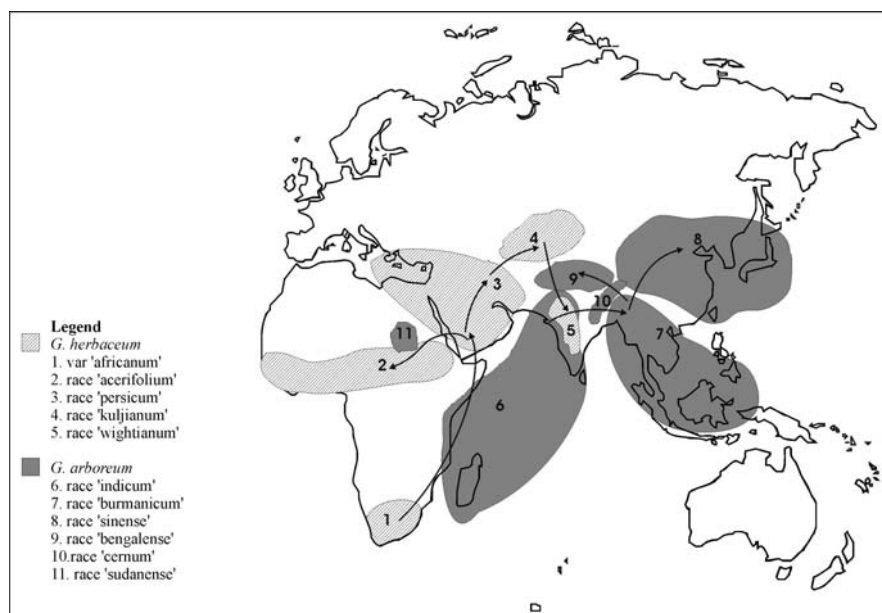


Fig. 1 Origin, domestication and spread of diploid cultivated cotton

North Africa. *G. herbaceum* also spread into Iraq, where the short summers and cold winters led to the development of the annualized race 'persicum'. This is thought to be the race that was introduced into Europe during the Alexander Era. In the short hot summers and long winters of central Asia, differentiation of early maturing race 'kuljianum' took place. Most recent introductions of annual forms in the intensive row-crop agriculture of the western peninsula of India gave rise to race 'wightianum'. Distributions of different races of *G. herbaceum* in Africa and Asia are presented in Table 1.

Table 1 Distribution of the main races of *G. herbaceum* in the Old World

Race	Region	Vernacular name/ remarks
persicum	Iran, Baluchistan, Afghanistan, Turkmenistan, Iraq and Mediterranean	Levant cotton
kuljianum	Western China and adjacent regions of erstwhile Russia	Not available
acerifolium	Arabia, Ethiopia, North Africa and Gambia	Commercially not important
wightianum	Western India and Persia	Waghad, Kalyana, Jayadhar
africanum	South-East Africa (Zimbabwe, Mozambique, Swaziland and the Transvaal lowveld)	Commercially not important

3.2 *Gossypium arboreum*

Despite no clear evidence of where *G. arboreum* was first domesticated, the Indus valley has been considered an important site of agronomic development and diffusion (Hutchinson 1954). The extensive races of *G. arboreum* have been elucidated by Silow (1944), Hutchinson (1954) and Brubaker, Bourland and Wendel (1999). *G. arboreum* race 'indicum' is the most primitive perennial form in western India (Fig. 1). The primitive land races of this race exist in Madagascar and coastal Tanzania, in addition to India. Dispersal of race 'indicum' eastward to northeastern India and Myanmar formed another perennial race 'burmanicum'. The western dispersal toward Egypt, Sudan and North Africa by Arabian traders formed race 'soudanense'. Introduction of race 'burmanicum' into the southern region of China developed into separate race 'sinense', an annual form of which was widely planted in the Yangtze and Yellow river valley (Guo, Zhou, Yang, Wang and Zhang 2006). Similarly, re-introduction of an annualized form of race 'burmanicum' into North India, where cultigens were exposed to frost led to the development of race 'bengalense'. These 'bengalense' cultivars are highly productive with coarse fiber. Annual types belonging to race 'cernuum' developed independently in the Assam hills of North-East India and Chittagang hills of Bangladesh. The cultivars belonging to race 'cernuum' are known as a genetic resource for bigger boll size, still cultivated in Gharo hills. Distribution of different *G. arboreum* races is presented in Table 2.

3.3 *Distribution of Diploid Cultivated Cotton in India*

In all the cotton growing zones of India, *G. arboreum* varieties and hybrids (involving lines of this species) are cultivated, however, *G. herbaceum* varieties are grown only in central and southern zones. Over the years of cultivation, various ecotypes have been identified in different zones. Bengal Desi (of race 'bengalense') in North India are highly productive, short (12–20 mm length) and very coarse (> 5.5 Micronaire value) with a prolonged maturity period (Fig. 2).

Table 2 Distribution of the main races of *G. arboreum* in the Old World

Race	Region	Vernacular name
indicum	Gujarat, Madagascar and coastal Tanzania	Rozi, Karungani, Gaorani, Mathio
burmanicum	Myanmar, parts of Bengal and Assam, Indochina, Malaysia and the East Indies	Dacca
cernuum	Assam hills, East-Bengal	Comilla
sinense	China, Manchuria, Korea, and Japan	
bengalense	Uttar Pradesh, Punjab, Sind, Central provinces of India	Bengaldeshi, Oomra
soudanense	Sudan regions of North Africa & West Africa	Senaar tree cotton

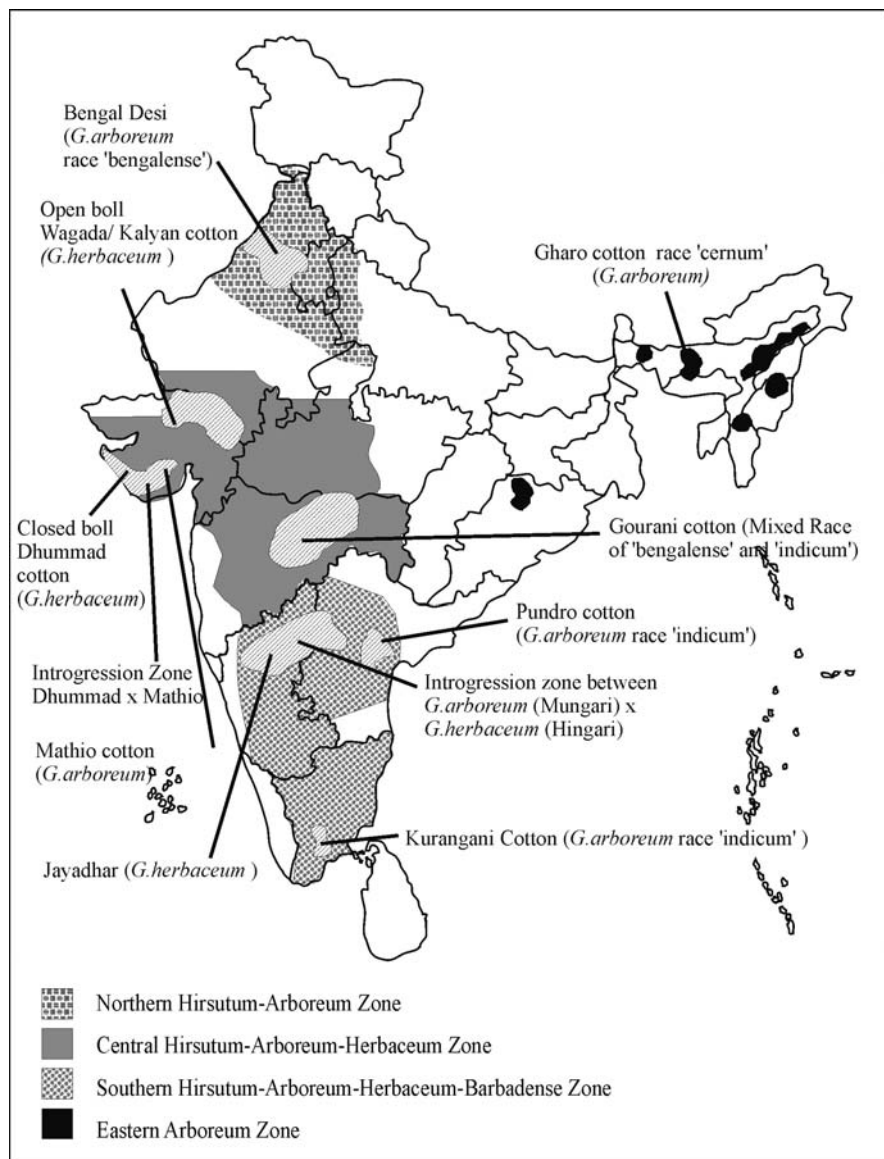


Fig. 2 Traditional diploid cotton growing areas *vis-a-vis* local types marked within three major cotton growing zones of India

Bengal Desi cotton varieties in Pakistan possess very high Micronaire (up to 11), the most-coarse cultivated cotton in the world. Mathio cotton grown under rainfed cultivation in Gujarat, Mungari cotton planted in early sown situations of Karnataka, Kurangani cotton grown in very harsh environmental

conditions of Tamil Nadu and Pundro cotton of Andhra Pradesh, known for its fine fibers, all belong to race 'indicum'. Sympatric cultivation of Mungari cotton in Karnataka and Mathio cotton in Gujarat, with *G. herbaceum* has resulted in a natural introgression zone, which was explored during 1999–2002 (Singh, Mohan, Kulkarni, Baitule, and Pathak 2003). Gaorani cotton with special adaptation to the shallow soil areas and rainfed environments of central India, grown in large scale, is a mixed type of races 'indicum' and 'bengalense'. Ghara cotton of race 'cernuum' are big boll types still cultivated in North-East zones of India and adjoining regions of Bangladesh.

The distribution of *G. herbaceum* cotton in India has been well presented by Narula, Acharya, Khadi and Kulkarni (2001). Dhummad cotton, a closed boll type (Fig. 3), cultivated under the coastal salinity of Gujarat (about 50,000 ha) is the most salt-tolerant diploid cultivated type in India (Babu, Punit Mohan, Singh and Singh 2003). Ginning of this cotton is most labor intensive, because the boll rind of cotton boll must be broken-open (usually using stone or iron rods) before passing the cotton to ginning machines (such ginning areas are called as Kaala Gins). Unlike Dhummad, Wagad/Kalyan types are open and semi-open types cultivated in rainfed ecology of Gujarat and parts of Rajasthan. Similarly the Jayadhar type in Karnataka is cultivated as either the sole crop or a mixed crop in onion and chili (about 125,000 ha) in residual soil moisture situation of the post rainy season. All these ecotypes of diploid cultivated cotton have short, coarse and weak fiber (except Pundro) and are used to blend with other cotton during spinning to achieve desired Micronaire values or length, and also for bed stuffing and non textile uses. The acreage of *G. herbaceum* has

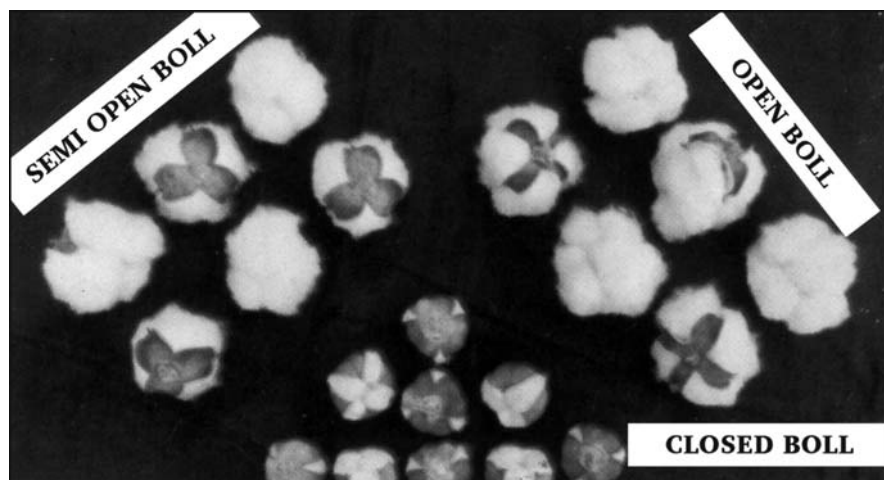


Fig. 3 Coverage of boll-rind after complete desiccation in closed boll (variety Dhummad), semi-open boll (variety G.Cot 17) and open boll type (variety Jaydhar) of *G. herbaceum* cotton

not been reduced in the last 50 years in India. On the contrary, the area of *G. arboreum* has shifted to cultivation of *G. hirsutum* and hybrids (Narula et al 2001).

4 Present Status of Diploid Cotton Cultivation

Old World Asiatic diploid cottons were of major economic importance during the early global expansion of commercial cotton production. Until the 1950's, major cotton growing areas of the Old World (Asia and Africa) were planted with cultigens belonging to either *G. herbaceum* or *G. arboreum*. With the introduction of New World cotton that had favorable plant type, high yield potential and superior fiber properties *vis-à-vis* overall economic value, the area under cultivation of diploid species was drastically reduced. During the 1940's and 1950's nearly 50% of cotton area of China and 90% of India was occupied by diploid cultivated cotton, which at present is 1% and 25% respectively.

Most of these cottons are cultivated in marginal, drought prone environments except *G. arboreum* race 'bengalense', grown in irrigated conditions of North India. Statistics presented by the International Cotton Advisory Committee, indicate the cultivation of diploid cotton in very limited scale (except India) in at least 10 countries spanning Asia and Africa (ICAC 1999; 2005). *Gossypium arboreum* is under cultivation in about 1.75 million hectares in seven countries, of South-East Asia or Southern Asia (Table 3). Namibia is the only African country growing *G. arboreum*, and in a very limited area. *Gossypium herbaceum* is planted in 0.77 million hectares in Iran, Turkey, India and Ethiopia. About 91% of the total *G. arboreum* acreages and 98% of *G. herbaceum* is in India, which has 94% of total diploid cotton acreage of the world.

Table 3 Area of diploid cultivated cotton in the world (ICAC 1999, ICAC 2005)

Country	Total cotton area (000 ha)	% area covered	Diploid cultivation cotton area (000 ha)*
<i>G. herbaceum</i>			
Ethiopia	113.1	1	1.13
India	9500	17	760
Iran	140	8	11.2
Turkey	150	4	6
<i>G. arboreum</i>			
Bangladesh	43.6	32.5	14.2
India	9500	17	1615
Namibia	3.3	15	0.5
Pakistan	3185	2	63.7
Thailand	36	3	1.08
Myanmar	270	21	56.7
China	4230	1	42.3

*Synthesized as per ICAC (1999) and ICAC (2005)

A synthesis by ICAC (2005) revealed that the global diploid cotton area is 13 times more than the Australian cotton area; 8 times more than Argentina and Egypt; 2.5 to 3.5 times more than Brazil and Turkey; and approximately 50% of the total cotton area of the USA and China. However, such a large cultivated area of diploid cotton growing regions are in general, either traditional diploid cotton growing tracts or areas mostly not suitable for cultivating tetraploids (*G. hirsutum* or *G. barbadense*), or drought/salinity affected lands. In the South-eastern region of Bangladesh, varieties/land races of *G. arboreum* (also called hill cotton) are a major component of traditional 'Jhum' cultivation, which is mixed crop cultivation (rice, sesame, cotton, maize, chili, pumpkin, cucumber etc) mainly on Chittagang hill slopes, where more than two crops are seeded and harvested separately. One cycle of cultivation involves 4–5 years. A major cultivation area of *G. arboreum* race 'bengalense' in Pakistan is in the Punjab and Baluchistan provinces where full season varieties are grown in low input conditions. Improved varieties or hybrids of *G. arboreum* are planted in high input conditions in the irrigated tracts of North India as well as in low input conditions in the central and southern parts of India. The land races *G. herbaceum* are traditionally grown in Khorasan and Gulestan regions of South East Iran under marginal conditions such as poor soil moisture and soil fertility, and hot wind. Farmers in such traditional tracts do not apply fertilizer or use improved management practices (Omran, Asadallah and Saeid 2007). Most of the produce from such a region is used for quilt making and as bed stuffing. For similar purposes, in the Aegean region of Turkey, land races of *G. herbaceum* are grown. Cultivation of 'bengaladesi' cotton, in Punjab and Sindh provinces of Pakistan is still practiced, and the produce from such cotton is very short and very coarse with Micronaire values touching double digits.

Land races of *G. arboreum* are grown in Southern regions of China, which are generally tall with longer growing periods, whereas varieties from North Eastern regions are generally shorter and with relatively abbreviated growing seasons. *G. herbaceum* grows only in Xinjiang and Gansu provinces, the north-western cotton growing regions.

5 Germplasm Resources

5.1 Status of Germplasm Collections

A global germplasm collection of diploid cultivated cotton has been maintained in the Central Institute of Cotton Research (CICR), Nagpur, India, which is the national center for collection, conservation, evaluation and documentation in the country. It holds the highest number of collections belonging to *G. arboreum* (1870) and *G. herbaceum* (530) (Table 4). Indian *G. arboreum* germplasm consists of native races such as 'cernuum', 'indicum' and 'bengalense', and some collections from China, which represent race 'sinense'. In addition to these,

Table 4 Gene banks of diploid cultivated cotton in the world

Country	Number of accessions		Place	Reference
	<i>G. arboreum</i>	<i>G. herbaceum</i>		
India	1870	530	Central Institute for Cotton Research, Nagpur, India	Anonymous (2005)
China	369	NA	Chinese Academy of Agriculture Science, Nanjing China	Liu <i>et.al.</i> (2006)
USA	1730	194	Southern Plains Agricultural Research Center (SPARC), Crop Germplasm Research Unit, College Station, Texas, USA	Anonymous (2005)
Iran	NA	42	Iran National Gene Bank and Iran Cotton Research Institute, Gorgan, Iran	Sheidal and Alishah (1998)
Vietnam	45	1	Nha Ho Cotton Research Center, Vietnam	Le (1996)
France	69	50	CIRAD, Montpellier, France	Dessauw and Hau (2006)

collections from Africa, probably belonging to race ‘soudanense’ and ‘comilla’ cotton of Bangladesh are also represented in Indian collections. Most of the *G. herbaceum* collections belong to race ‘wightianum’ and very few to var *africanum*. Systematic race-wise characterization of germplasm has been very difficult due to overlapping morphological traits, and non-availability of race-specific phenotypic and molecular markers. However, such a race-wise classification is yet to be done in India using morphological and molecular markers. The majority of accessions of both species are characterized for different agronomic and fiber properties, and using such data, germplasm groupings have been made to help the breeders (Singh, Baitule and Pathak 2001). Elite types with specific traits such as red plant body, brown lint, nectariless, spotless flower, big bolls, higher ginning turnout, early maturity, high seed oil and long staple have been identified from germplasm accessions (Singh et al 2001; Singh, Khadi, Kulkarni, Mohan and Anjali 2004).

The majority of diploid A-genome accessions of the USA are imported from India and China during 1994–95 (Anonymous 2005). The French cotton germplasm collection, preserved in Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) in Montpellier (France) has variability sampled from 8 expeditions in African countries and the Arabian peninsula. The collections include *G. arboreum* race ‘soudanense’(22), *G. herbaceum* var *africanum* (2), cultivars of both species and other ancestral forms (18 in *G. herbaceum* and 17 in *G. arboreum*) that are yet to be characterized (Dessauw and Hau 2006). The *G. herbaceum* germplasm of Iran consists of 42 ecotypes collected at the altitude range of 10–500 m, that are well characterized morphologically and cytologically. Xiang (1988) collected 369 samples of

G. arboreum race 'sinense' from 20 provinces and characterized using 72 characters that showed tall plant stature with longer growing periods of the southern region accessions compared to the short height and early-maturity of the north-eastern region. Vietnamese cotton collections are predominantly *G. arboreum* race 'burmanicum'. Although collections of wild and cultivated diploid cottons exists with Sudan at Shambat, detailed information about this collection is not available (Ali, Ahmed, Misaka, Latif, Elsiddig and Babiker 2003).

The table showing germplasm status is incomplete due to a paucity of information from Pakistan, Bangladesh, Thailand, and other South-East Asian countries, where domestication of *G. arboreum* occurred; and erstwhile Russian republics adjacent to western China, where domestication of *G. herbaceum* race 'kuljianum' took place.

5.2 Germplasm Expedition

Since 1979, CICR, along with the National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India conducted 14 explorations in six regions of diversity in India, which were mainly aimed at sampling the rich variability in diploid cultivated cotton (Singh, Mohan, Kulkarni, Baitule and Pathak 2003). The north-east region of India, the home of *G. arboreum* race 'cernuum' was explored four times since 1979 (Table 5). Similarly, the southern coastal region including Tamil Nadu and Andhra Pradesh was also explored 4 times, mainly to collect *G. arboreum* race 'indicum' and *G. herbaceum* types. In the recent explorations, 'cernuum' cotton accessions with up to 7.3 g boll weight, burst capsule length of 17.5 cm and lint% up to 46, 'indicum' accessions with long (34 mm halo length) and fine fibers, and Dhummad cotton accessions from southern coast of Gujarat having closed bolls with tolerance to high levels of salinity, were collected along with accessions having other diverse morphological features (Singh et al 2003).

The last expedition held by CIRAD was in 1988 in the African center of diversity and by China in 1988 in Chinese cotton growing regions (Xiang 1988; Dessauw and Hau 2006; Guo et al 2006). Exploration trips outside of India, aiming to collect diploid cultivated cotton are not known to have been carried out in the past 15-20 years. However, gene banks in the countries of South-East Asia hold local collections in limited numbers from the respective countries.

Recently, an expedition was carried out in the Egadi archipelago (Italy) for safeguarding of crop genetic resources on Italian minor islands (Sicily), where two very rare accessions of cotton (called 'u cuttuni' and 'mattola') collected, merit a special note (Laghetti, Hammer, Cifarelli, Branca, Diederchsen and Perrino 2002). During the ninth century Arabs introduced cotton (*G. herbaceum*) to Sicily. On Favignana (a region in Sicily), the ancient and traditional cultivation of cotton (particularly resistant to salt water and sometimes used locally for forage before its flowering) ended completely in 1950. However, two

Table 5 Regions explored, material collected and characters identified through CICR-NBPGR expeditions in India

Region Explored	Year	Species	No. of accessions.	Character
NIEH Region - mainly Assam, Meghalaya, Nagaland, Manipur, Tripura	1979, 1984, 1990, 2000	<i>G. arboreum</i> race 'cernuum' and perennials	223	Big long bolls, high boll weight, high ginning out turn, high seed number, coarse fiber, short staple and high locule retentivity
South Coastal Region (TamilNadu and Andhrapradesh Coasts)	1979, 1981, 1993, 2002	<i>G. arboreum</i> introgressed 'indicum' and <i>G. herbaceum</i> perennials (feral) and very few <i>G. hirsutum</i> ,	116	Small bolls, medium staple, salt and drought resistant/tolerant types, bourbon types (<i>G. hirsutum</i>) drought tolerant, early maturity
Gujarat (Kachchh Saurashtra Region)	1979, 2000	<i>G. herbaceum</i>	50	Round and closed bolls, high ginning turn out and drought tolerant
East Coast Tract (Orissa)	1989	Mostly <i>G. arboreum</i> , some <i>G. hirsutum</i> , perennials and introgressed types	54	Small bolls, high ginning outturn
Jammu Region	1991-92	<i>G. arboreum</i>	45	High elevation and cold tolerant types
Karnataka	1979, 2001	<i>G. arboreum</i> , <i>G. herbaceum</i> and very few <i>G. hirsutum</i>	76	High boll number and high ginning outturn, robust plant, smooth fibre, compact plants, early maturity, dwarf plants, tall plants

farmers who still grow plants belonging to the old local varieties for family uses and for ornamental purposes were discovered during the expedition. In 1985, (Hammer et al. 1986) rare relic plants of cotton used for local production of fibre and as medicinal plants were collected from some Sicilian home gardens. These findings and the history of cotton in south Italy suggest that a specific collection mission should be arranged to find and safeguard the unique Sicilian cotton germplasm before its extinction.

Regional diversity is known to exist among diploid cultivated cotton, but is poorly represented in the global collections especially with respect to *G. herbaceum* species from South Africa (var '*africanum*', progenitor of tetraploid cotton), the African Sahel (race '*acerifolium*'), and the countries of Asia Minor (race '*persicum*' in Turkey, Iraq, Iran and Pakistan). Similarly, variability of *G. arboreum* has been poorly represented from South-East Asian countries such as Nepal, Bangladesh, Myanmar, Thailand, Laos, Vietnam, Cambodia, Malaysia, and Indonesia (races '*burmanicum*' and '*cernuum*') that exist in small family plots, in cultivation in small pockets of certain regions, or as dooryard plants. Accessions from these places are thought to be important sources of biotic and abiotic stress tolerance. These facts accentuate the need for regional surveys and collections of germplasm through joint expeditions from the above-mentioned regions before these cotton genotypes go extinct.

6 Genetic Diversity

Gossypium herbaceum and *G. arboreum* cotton breeding is handicapped by the paucity of information on genetic diversity. A systematic genetic assessment of the gene sources will help to reduce redundancy and construct core germplasm collections that will enhance utilization of germplasm. All cultivated diploid germplasm in India has been evaluated at least for one year, and some sets for more than one year for economic characters (Singh 2007, CICR, Nagpur, India personal communication). Fiber properties of about 50% of the collections have been recorded and a similar fraction of accessions have been evaluated for disease resistance. Less than 40% of the accessions were screened for tolerance to insects. A set of germplasm has been classified into 22 trait-specific groups based on 3–4 years of data on plant, boll, seed and fiber characters (Singh et al 2001).

Data processing is being done at CICR, Nagpur, India, to assess genetic diversity using different tools, to generate a core collection, and also to make trait-specific classifications of germplasm. In the germplasm accessions preserved in the USA, significant genetic variability for 41 characters in both *G. herbaceum* (84 accessions) and *G. arboreum* (272 accessions) was observed with overlapping clusters due to no clear separation of ranges of many morphological characters between the two species (Stanton, Stewart, Percival and Wendel 1994). In Iran, 42 ecotypes were grouped into five clusters based on six

morphological characters (Sheidai and Alishah 1998) and into four groups based on karyological observations such as variation of chromosomal size and volume, satellite chromosome numbers and karyotype formulae (Omran et al 2007). The clustering based on morphometric traits was almost similar to that of karyological studies. A high level of genetic variation (revealed by 358 simple sequence repeat (SSR) markers) was observed among the *G. arboreum* accessions collected from different regions of China, with positive significant correlations between cluster groupings and geographic distances (Liu et al 2006). Genetic variability studies using SSR markers in *G. arboreum* race 'sinense' collected from different provinces of China revealed higher variability in land races of the southern region than from the Yangtze and Yellow river valley regions (Guo et al 2006). Indian varieties of *G. arboreum* and *G. herbaceum* formed two distinct clusters when genotyped using RAPD and AFLP markers (Rana and Bhat 2004). Systematic morphological and molecular characterization of the diploid cultivated cotton gene pool is essential to enhance its utility in both diploid and tetraploid cotton breeding.

7 Relationship with Species of Other Genomes

The relationship within the *Gossypium spp* has been well enumerated by Belgian authors (in French) Ndungo, Demol and Marechal (1988) based on the assembled cytological data of hybrids between different species. Based on this compilation, a gene pool concept of diploid cultivated cotton has been described by Khadi, Kulkarni, Manjula and Katageri (2002). Cytological data of crosses involving at least one A-genome species was used to establish the relationship with other genomes (Table 6).

Both A- genome diploid species are biological, differing for one translocation and forming 11 to 13 bivalents during meiosis. African B-genome species are close to A-genome species and produce fertile hybrids. Average bivalent frequency during meiosis in A x B crosses ranged from 10 to 13. The A-genome

Table 6 Meiotic chromosomal configurations in different inter-specific hybrids involving the A-genome

Hybrid	Meiotic chromosomal configuration (range)			
	I	II	III	IV
A x A	0 – 0.22	11.13 – 13.0	–	0 – 0.93
A x B	1.2 – 5.4	9.96 – 12.13	0 – 0.26	0 – 1.22
A x C	2.8 – 21.6	2.2 – 9.8	0 – 0.65	0 – 0.38
A x D	9.6 – 23.6	2.43 – 7.8	0.01 – 0.19	0.04 – 0.12
A x E	11.7 – 24.1	0.95 – 7.13	0 – 0.2	–
A x G	16	5	–	–
AD x A	13	2 – 13	–	1 – 9

Abstracted from compilation made by Ndungo, Demol and Marechal (1988)

and B-genome serve as primary gene pools for diploid cultivated cotton improvement. Fertile and highly heterotic plants of crosses between *G. herbaceum* and *G. anomalum* have been obtained (Kulkarni, Khadi, Bisankoppa, Hussain and Narayanan 2004). Diploid A- and tetraploid AD crosses produce sterile plants due to their triploid nature, although these two are related to each other due to the A-genomic content in the AD tetraploid species. Higher average univalent frequency observed during the meiosis of hybrid plants between A-genome and American D- genome diploids, and Afro-Arabian E-genome suggest that these are distant from A-genome. Similar is the case with Australian species belonging to C- and G-genome that produce sterile hybrids when crossed with A-genome species. Hybrids of A- and G-genome species (*G. australe* Mueller) are sterile (Kulkarni, Khadi and Sangam 2001).

8 Breeding Cultivated Diploids

Until the 1950's, Old World cotton varieties of *G. herbaceum* and *G. arboreum* were grown in major parts of the Asian subcontinent reaching from Turkey to Korea. *Gossypium hirsutum* varieties and/or hybrids (in the case of India) have replaced *G. arboreum* and *G. herbaceum* to a great extent in many countries. The major causes for this change were due to fiber, boll and plant features of *G. arboreum* and *G. herbaceum*. Diploid cultivated cottons are characterized by short (<23 mm fiber length), coarse (>5.0 Micronaire) and weak (< 20 g/ tex at 3.2 mm gauge) fibers, which were not suitable for spinning by mechanized textile technology. The lanky plant stature with indeterminate growth and small bolls with less locule retention are major hurdles in cotton harvesting, impairing the economics of cultivation. Diploid cotton varieties became obsolete in most parts of Asia, and genetic improvement gradually ended in many countries, except India.

Although diploid cultivated cottons are grown in 10 countries, organized breeding efforts are in place in India alone. In Pakistan and Bangladesh breeding of diploid cotton has almost ceased. These species are still being used for various genetic and biotechnological studies in many laboratories around the world. *G. arboreum* and *G. herbaceum* are being used as donor species in introgressive breeding to improve tetraploid cotton, especially for disease resistance and insect tolerance (Ansingkar, Khadke, Borikar and Bhosle 2004, Kulkarni 2002). In India, *G. arboreum* and *G. herbaceum* genetic improvement has been part of the All India Coordinated Cotton Improvement Project (AIC-CIP), which is supported by special projects in Technology Mission on Cotton (TMC), Mini Mission I and Rainfed Cotton Agro-ecosystem (RCPS) Programs of the National Agriculture Technology Project (www.cicr.nic.in). The development of genetic materials, evaluation in systematic trials and recommendation of elite genotypes for cultivation has been well structured in India. As per the recent AICCIP reports, about 25 genetically improved elite genotypes of *G. arboreum*

are evaluated every year in a national trial (Breeding trial 22) in about 15 locations across all three cotton growing zones of India both in irrigated and rainfed agro-ecosystems for various economic traits. Similarly, elite *G. herbaceum* genotypes (10-15) are evaluated (Breeding trial 34) in four to five locations in southern and central India only in rainfed conditions. In addition to these, about 15 hybrids belonging to both intra-specific (*G. arboreum* x *G. arboreum*) and inter-specific (*G. arboreum* x *G. herbaceum*) groups, developed using genetic male sterility (GMS), are evaluated in national trials (Breeding trial 25) conducted in about 15 locations across three cotton growing zones in both irrigated and rainfed conditions. Results of these trials provide hybrid and/or variety options to the farmers of India for planting diploid cultivated cotton.

8.1 Diploid Cotton Breeding Centers and Breeding Objectives

Genetic improvement for various economically important traits in both diploid cultivated cottons is focused in special projects of TMC and in AICCIP operating in all three cotton growing zones of India. Four major research centers (Ludhiana, Hissar, Sriganaga Nagar and Kanpur) in the North work in developing high yielding *G. arboreum* varieties adapted to irrigated agro-ecosystems. To date, varieties developed by these centers generally have coarse and short fiber. In addition to high yield under favorable conditions, fiber quality improvement, especially fiber length (25–27 mm) and fineness (4–5 micronaire), are priorities in breeding. Inducing earliness for double cropping systems, synchronous flowering to reduce the number of cotton hand pickings (harvesting), manipulating plant habit to bring erect plant stature, improving tolerance to bollworms, and increasing boll weight and locule retention to reduce harvesting losses, are also researchable issues being considered by breeders.

Gossypium arboreum varieties developed by research stations in southern (Dharwad, Nandyal, Kovilpatti, Guntur) and central (Parbhani, Nagpur, Akola and Surat) India are adapted to rainfed agro-ecosystems and generally have slightly longer (25–26 mm) and finer fiber (4.5–5 Micronaire) than North Indian varieties. Improving genetic yield potential in rainfed cultivation has been the first priority where breeders are exploiting the drought tolerance of *G. arboreum* ((Deshpande, Narula and Baig 2001), followed by improving fiber quality similar to *G. hirsutum*. Erect plant type, large boll size, synchronous flowering, and locule retention are plant related traits being targeted for genetic improvement. Breeding for resistance to fungal diseases and bollworms is very important to reduce losses in all the cotton growing zones.

Development of high yielding *G. herbaceum* varieties has been a priority only in federal states Gujarat (Surat, Bharuch) and Karnataka (Dharwad), where it is predominantly grown in rainfed-agro-ecosystems and residual moisture conditions of the post rainy season, respectively. Transferring good fiber quality traits, and earliness from either *G. arboreum* or *G. hirsutum* has been a necessity

for the mills and farmers. Recently, development of hybrids with wide adaptability has been a priority. Public sector breeding institutes, and private seed companies have been engaged in commercial hybrid development using genetic male sterility (GMS) in various parts of India that mainly aim to exploit of F_1 heterosis available for various traits.

8.2 Pattern of Variability

Exploitable genetic variability was observed for yield and yield components in *G. arboreum* (Singh and Singh 1984; Kumar and Rajmani 1994) and in *G. herbaceum* (Singh 1983), and for fiber properties in *G. arboreum* (Singh 1986) and *G. herbaceum* (Singh 1983). The exploitable variability for various characters within each species indicate potential for genetic improvement both in *G. arboreum* and *G. herbaceum* (Tables 7 and 8). Additive genetic variance generally observed for boll weight, lint% and fiber length would provide opportunities to the breeders to use genetic variability for these traits. Prospects of improvement of boll weight, fiber length and lint% in *G. arboreum* are much higher than *G. herbaceum*, because of the availability of genetic stocks for the respective characters. Germplasm accessions with big boll belonging to race ‘cernuum’ (Garro hill cotton), long fiber of race ‘indicum’ and higher lint% of

Table 7 Potential for improvement of different characters in the germplasm of *G. arboreum*

Character	Potential*	Achieved	Gap
Boll weight (g)	6.50	2.90	3.60
Lint (%)	55	41	14
2.5% span length (mm)	30	25	5
Earliness (days)	135	180	45
Harvest index (%)	58	25	33
Seed oil (%)	24.5	20	4.5
Seed cotton yield plant ⁻¹ (g)	170	90	80

* Highest value observed in the working collection

Table 8 Potential for improvement of different characters in the germplasm of *G. herbaceum*

Character	Potential*	Achieved	Gap
Boll weight (g)	2.8	2.0	0.8
Lint (%)	43	36	7
2.5% span length (mm)	26.9	23.0	3.9
Fiber fineness (Micronaire)	2.4	3.4	1.0
Fiber maturity (%)	95	69	36
Uniformity ratio (%)	52	49	3
Seed oil content (%)	20.4	16.5	3.9
Seed cotton yield plant ⁻¹ (g)	109	40	69

* Highest value observed in the working collection

race ‘bengalense’ could be used in elevating the trait values in elite agronomic backgrounds. A preponderance of non-additive genetic variation for seed cotton yield, boll number and seed oil% might prompt breeders to use heterosis breeding for better utilization of variability. Although several studies (mostly using accessions/lines of the same races) have given clues to the nature of genetic variability available for important characters in different materials, there is good scope to unearth further information with systematic experiments involving diverse materials. A wide range of genetic variability in *G. herbaceum* genetic stocks observed in erstwhile USSR, increases the potential for genetic improvement as large variability has been reported for boll weight (1.5 – 5.0 g), mean fiber length (18-28 mm) and lint% (20 – 35%) (Singh 1998).

8.3 Breeding Approaches

Various breeding methods have been used to exploit genetic variation and variety development in diploid cultivated cottons. Of the 66 varieties developed for cultivation in India, 54 are *G. arboreum* and 12 are *G. herbaceum* (Table 9). Pedigree breeding was predominantly used in variety development, as about 40% of the released cultivars were developed through selection in segregating populations generated from single crosses of two parents of the same race. This warrants utilization of inter-racial crosses to combine different traits to achieve higher yields and good quality. Mass selection has been used in the development of 16 *G. arboreum* varieties, among which 6 were selected from a Gaorani bulk or lines, which had wide adaptability and high yield potential.

Table 9 Number of *G. arboreum* and *G. herbaceum* varieties developed using different breeding methods (modified after Anonymous 2006)

Breeding Method	<i>G. arboreum</i>	<i>G. herbaceum</i>
Land races	2	1
Mass selection	15	–
Mutation breeding	1	1
Backcross breeding	–	1
Pedigree selection		
i) Single cross		
a. Inter- racial	1	–
b. intra-racial	20	5
c. inter-specific	1	1
ii) Double cross	2	–
iii) Three way cross	4	2
iv) Multiple cross		
a. Inter- racial	3	–
b. Intra- racial	5	1
Heterosis breeding	10*	

* five each of inter-specific (*G. arboreum* x *G. herbaceum*) and intra-arboreum



Fig. 4 Dhummad variety of *G. herbaceum* having closed bolls cultivated in coastal salinity of Gujarat, India (See Color Insert)

Among the 10 hybrids released for cultivation four were developed based on conventional emasculation and pollination, and the remainder using GMS. Many private seed companies are trying to exploit hybrid vigor using GMS technology and some are marketing productive hybrids. A closed boll type land race of *G. herbaceum*, Dhummad (Fig. 4), has been cultivated for six decades in coastal saline areas of Gujarat, and another open boll type *G. herbaceum* variety, Jayadhar, developed through inter-specific hybridization (*G. herbaceum* x *G. arboreum*) and selection has been cultivated for five decades in South India. In our experience, inter-specific hybridization involving *G. arboreum* and *G. herbaceum* results in slow gain from selection due to abnormal segregation and mortality in F_2 , where fewer than 10% of plants are productive requiring breeders to plant higher population size. Such plants in F_2 will not breed true in later generation. Hence, we recommend use of either a backcrossing or biparental mating in segregating generation or single seed descent method for achieving relatively better genetic gains. Utilization of germplasm for genetic improvement of yield, boll and fiber properties have been described below.

8.3.1 Seed Cotton Yield Improvement

Genetic gains for seed cotton yields in India were predominantly through varietal development, which exploited additive genetic variation. Recently, exploitation of non-additive genetic variation was achieved through hybrid

development. Genetic gains for seed cotton yield were evident over time, as varieties released during the 1960s and early 1980s had less seed cotton yield potential (1600 to 1800 kg ha⁻¹), compared to the varieties such as LD 327, HD 107 and RG 8 (2500–2600 kg ha⁻¹) and a hybrid AAH 1 (3900 kg ha⁻¹) in north Indian irrigated conditions (Sandhu 1989; Singh 1998; Lather, Chhabra, Sangwan and Siwach 2001). Although seed cotton yields of recently developed varieties adapted to rainfed ecosystems of south and central India is high (500–600 kg ha⁻¹ of 1980s to 800–1000 kg ha⁻¹ of 2000s), it is low compared to North India. Higher genetic gains in the recent varietal development efforts could be attributed to stabilization of potential recombinants from the inter-racial crosses involving ‘bengalense’ and ‘cernuum’ (north Indian irrigated tracts) and ‘bengalense’, ‘indicum’, and ‘cernuum’ (South and central India). No significant gains have been observed in lint percentage, which remains at 35–41%. *G. herbaceum* varietal development has resulted in the development of semi-open or open boll types that contributed to reducing harvesting losses especially in Gujarat (Waghad/Kalyan and part of Dhummad tracts) in addition to higher yields.

Seed cotton yield improvement through the hybrid route in *G. hirsutum* and *G. barbadense* prompted diploid cotton breeders to try out the using *G. herbaceum* and *G. arboreum* same in the early 1980s. Positive heterosis has been observed for many traits over the best parent (Khadi and Kulkarni 1999). Hybrids produced through conventional hand emasculation and pollination (Doak, 1934; Mehta and Patel 1983) showed heterosis of up to 185% for seed cotton yield (Singh, Kandola and Nagi 1975; Batade 1983; Rajput, Meshram, Kalpande, Golhar and Bharad 1998). Conventional hybrid seed production was not successful, because of delicate flower structure leading to low frequencies of crossed boll setting; and lower seed index leading to less hybrid seed yield. Alternatively, GMS technique was used for exploitation of heterosis. A single recessive gene is involved in the expression of GMS. Of the two sources of GMS, Hissar source (DS₅) was isolated from a spontaneous mutation in *G. arboreum* and has small white flowers (Singh, Rameshkumar and Lather 1992). Akola source (GAK 423A) was developed by transferring genome of *G. arboreum* into *G. anomalum* cytoplasm and has large yellow flowers (Meshram and Wadodkar 1992). Thermo-sensitive male sterility is being investigated at CICR (Nagpur, India), which might further ease parental line maintenance and hybrid development. Several hybrids have been developed using GMS sources by both public and private sectors that have provided alternative cultivar options for the farmers. In some inter-specific hybrids, *G. herbaceum* has been used as one of the parents to provide better adaptation of hybrid to the adverse climatic conditions of southern and central India.

8.3.2 Boll Size and Fiber Quality Improvement

Efforts in fiber quality improvement have been very effective in southern and central India, which mostly utilized ‘indicum’ source for long and fine fiber.

Reports of mass selection in this race have generated a wide range of exploitable genetic variation for fiber properties (Rao, Deshpande and Khadi 2004), but boll size remained small. Although 'cernuum' race was extensively used to improve boll size, it contributed undesirable fiber properties. These conflicting problems were addressed through *G. hirsutum* gene introgression that was efficient in genetic enhancement of fiber properties and boll features of *G. arboreum* (Kulkarni, Khadi, Deshpande, Sreenivasan and Bisankopp 2003). This process was called as 'Hirsutisation' of *G. arboreum* cotton and the lines developed through this process are called 'hirsutised *G. arboreum*'. Using colchicine, *G. arboreum* chromosomes were doubled making it an auto-tetraploid ($A_2A_2A_2A_2$), which was crossed to *G. hirsutum* ($A_1A_1D_5D_5$). As the F_1 embryo was abortive, pollen having 27.6% fertility was used to backcross *G. arboreum* and advanced generation (F_6) progenies (68) were evaluated across India. Results clearly showed genetic improvement with respect to fiber qualities in addition to yield and boll features (Table 10). Genetically diverse progenies with equivalent fiber quality of *G. hirsutum* (Fig. 5) with high yield potential were evaluated in different environments and two varieties (PA 402 and DLSA 17) were identified for cultivation in peninsular India under rainfed ecosystems (Kulkarni et al 2003; Deshpande, Baig, Kulkarni, Kakde and Matishchandra 2004). Although not high, relatively better genetic gain was noticed for boll size (Fig. 5) and seed index (Fig. 6) due to *G. hirsutum* introgression. These improved characters in *G. arboreum* would contribute to heterosis breeding, where seed and boll size are major constraints in seed production, and increased fiber length is highly desired for commercial production. Efforts to improve *G. herbaceum* through introgression breeding using *G. hirsutum* has produced *G. herbaceum* progenies with fiber length up to 30 mm and boll weight to 3.0 g (Manjula, Khadi, Pawar, Immadi and Katageri 2004). *G. anomalum* has been found to be a good source to improve the strength and fineness of *G. herbaceum* (Kulkarni, Khadi, Bisankopp, Hussain and Narayan an 2004). Backcross progenies aimed to derive *G. herbaceum* pure lines of inter-genomic cross (*G. herbaceum* x *G. anomalum*) showed a wide genetic variability for strength (19–43 g/tex at 3.2 mm gauge). In the TMC projects improvement of fiber properties of *G. herbaceum* is underway using hirsutised long staple *G. arboreum* varieties especially DLSA 17.

Shedding or shattering of locules from opened bolls is a serious problem in most of the varieties of *G. arboreum* especially in north India. Garo hill cotton belonging to race 'cernuum' is a good source of locule retention (Singh and Nandeshwar 1983). Utilizing this source, Singh and Narayanan (1987) have developed NAS 3, NAS 4 and NAS 5 that combine high yield potential of race 'bengalense'. CINA 36, an improved race 'bengalense' line, has been registered for good locule retention (www.cicr.nic.in). The force (in grams) required to pull the seed cottons from fully opened bolls was measured while breeding improved lines (Singh and Narayanan 1987)

Table 10 Gross maximum genetic gain observed in the *G. arboreum* progenies (F₆) obtained through introgression of *G. hirsutum* and *G. arboreum* for yield and fiber quality characters across cotton growing zones of India

Character	Parameter	Dharwad (South)	Parbhani (Central)	Delhi (North)	Mean
Boll weight (g)	Maximum	3.3 (44)*	3.3 (47)	3.1 (58)	3.1 (49)
	Local check	2.1	2.1	1.8	2.1
	% increase	57.1	57.1	72.2	47.6
Seed cotton yield (kg ha ⁻¹)	Maximum	1800 (6)	1174 (9)	2928 (0)	1766 (8)
	Local check	1001	774	2603	1459
	% increase	79.8	51.7	4.5	21.0
Ginning turn out (%)	Maximum	42.4 (6)	40.1 (2)	42.2 (2)	41.6 (8)
	Local check	35.5	37.2	37.8	36.8
	% increase	19.4	7.8	11.6	13.0
Seed index (g)	Maximum	10.5 (15)	9.3 (43)	8.3 (66)	9.4 (61)
	Local check	7.3	6.3	5.1	6.2
	% increase	43.8	47.6	62.7	51.6
2.5 % span length (mm)	Maximum	28 (49)	27.1 (8)	28.3 (49)	27.4 (41)
	Local check	23.2	24.5	18.3	22
	% increase	20.7	10.6	54.6	24.5
Tenacity (g/tex) at 3.2 mm gauge [§]	Maximum	25.2 (50)	24.9 (40)	25.9 (34)	25.2 (50)
	Local check	19.7	19.7	16.9	18.8
	% increase	27.9	26.4	53.3	34.0
Micronaire [§] value	Maximum	4.1 (13)	3.8 (45)	4.2 (6)	4.2 (14)
	Local check	5.3	5.2	7.0	5.8
	% increase	-22.6	-26.9	-40.0	-27.6
Uniformity ratio (%)	Maximum	52.5 (0)	53.3 (18)	53 (6)	51.9 (1)
	Local check	48.5	46.8	50.5	48.6
	% increase	8.2	13.9	5.0	6.8
Elongation (%)	Maximum	7.9 (8)	7.5 (64)	8.8 (3)	7.7 (8)
	Local check	6.2	5.4	6.4	6.0
	% increase	27.4	38.9	37.5	28.3

* Figures in parenthesis are the number of progenies significantly superior to high value check, irrespective of *G. arboreum* and or *G. hirsutum*, except for traits, boll weight, seed index and 2.5 % span length where figures in parenthesis are number of progenies significantly superior over *G. arboreum* (LC); § Fiber strength expressed in ICC mode; Micronaire value = 4.5 or less is desirable.

8.3.3 Disease Resistance and Insect Tolerance

Grey mildew (*Ramularia areola* Atk), Leaf spot (*Alternaria macrospora*) and Fusarium wilt are important diseases and bollworms (*Helicoverpa armigera*, *Earias vitella*, and *Pectinophora gossypiella*) are insect pests causing economic damage. Stringent screening of materials right from germplasm accessions to advanced generation lines in a wilt nursery facility under AICCIP made it possible to incorporate resistance into all the released varieties of *G. arboreum* and *G. herbaceum*. Grey mildew disease causes defoliation of the entire plant under severe infestation that results in 30% to 60% yield loss (based on the stage of infestation). Two germplasm accessions of race 'cernuum' (30805 and

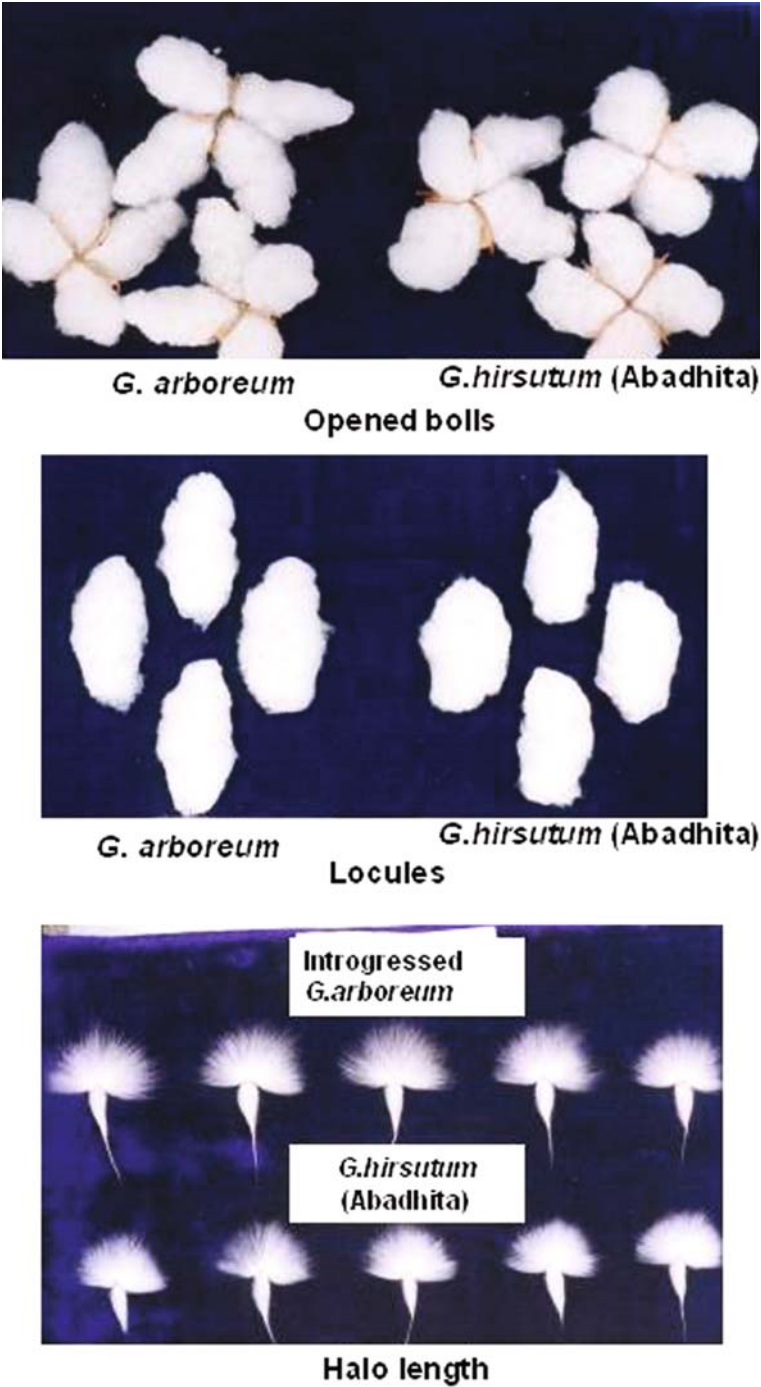


Fig. 5 Comparison of introgressed *G. arboreum* (backcross derivatives of cross between *G. arboreum* and *G. hirsutum*) with *G. hirsutum* for (a) boll size (b) locule size and (c) halo length (See Color Insert)



Fig. 6 Improved seed size in introgressed *G. arboreum* lines (AH 56 and AH 29) in comparison with *G. arboreum* parent (PA 140), *G. hirsutum* parent (Poornima) and small seeded *G. arboreum* line (See Color Insert)

30838) and one of race 'bengalense' (G 135-49) have been registered for immune reaction to grey mildew. These accessions serve as superior materials for developing populations to identify markers closely linked to disease resistance. Although *G. arboreum* are more tolerant to bollworms than *G. hirsutum*, during severe outbreaks economic losses are observed. As there is no germplasm

accession that provides durable resistance to bollworms, the transgenic approach of transferring an insecticidal protein *Cry1AC* gene into long staple hirsutised *G. arboreum* (DLSA 17 and PA 402) is in progress to develop bollworm resistant genotypes (Chinchane, Nandeswhar, Deshpande and Chinchane 2004). Diploid cultivated cotton genotypes are sources of resistance to the sucking pest complex (hoppers, thrips, mites and whiteflies) and immune to leaf curl virus disease (Mehethre, Shinde and Pardeshi 2004) and are being used in genetic improvement of *G. hirsutum* (Kulkarni 2002; Ansingkar et al 2004). In addition to pubescence, the presence of a lower palisade layer, relatively higher distance between lower epidermis of midrib and phloem, and densely arranged midrib cortex cells are reasons for sucking pest resistance, which were transferred to *G. hirsutum* through inter-specific hybridization (Kulkarni 2002; Ansingkar et al 2004).

9 Genetic Materials for Biotechnological Studies

Backcross derivatives of the cross between 4n *G. arboreum* ($A_2A_2A_2A_2$) and *G. hirsutum* ($A_1A_1D_5D_5$) are highly suited (Kulkarni et al 2003; Khadi, Kulkarni, Bisankopp, Vamadevaih 2004) for genetic studies, because such a cross facilitates isolating recombinants between A_2 genome of *G. arboreum* and highly evolved differential A_1 (*G. herbaceum*) genome of AD tetraploid, which otherwise are not possible to obtain. It also provides great scope for preferential pairing of A and D chromosomes in F_1 ($A_2A_2A_1D_5$) and such recombination was noticed in our study as 1 to 3 quadrivalents were observed in the pollen mother cells of F_1 plants (Deshpande, Kokate, Kulkarni and Nerkar 1992). Recombinant progenies from such a cross also pave the way for studying the role of D-genome specific genes in fiber quality improvement (Khadi et al 2004). The cross between *G. herbaceum* (AA) and *G. anomalum* (BB) resulted in an F_1 with strongest fiber (51.8 g/tex at 3.2 mm gauge) (Kulkarni et al 2004) and back cross derivatives of such a cross might be suitable for mapping the primary determinants of lint fiber evolution. This provides the genetic stocks that are necessary to identify the genetic map locations of the specific mutations responsible for the development of spinnable fibers, using an established genetic map (Rong et al 2004). *G. herbaceum* could be crossed with *G. australe* through conventional hand emasculation and pollination and the F_1 could be used for backcrossing with *G. herbaceum* to isolate plants with glandless seed and glanded plants in addition to elucidating the relationship between A and G genomes (Kulkarni et al 2001).

Acknowledgments We are grateful to Miss C. Madhura for help in information collection and tables, and Miss Shreya Kulkarni for proofreading. Our thanks are also to funding agencies, National Agricultural Technology Project and Technology Mission on Cotton and Mini Mission I of the Indian Council of Agricultural Research, New Delhi, India.

References

- Ali, A. M., Ahmed, O.M., Misaka, B.C., Latif, A.H., Elsiddig, K. and Babiker, E.A. (2003) Characterization of cotton germplasm and its utilization in breeding for major production constraints in Sudan. In: A. Swanepoel(Eds), Proceedings of the World Cotton Research Conference-3 on "Cotton Production for the New Millennium", 9–13 March 2003, Cape Town, South Africa, pp. 18–21.
- Anonymous (1999). Indian cottons over the years-A Compendium of Technological Properties. Central Institute for Research on Cotton Technology Mumbai. 16–143.
- Anonymous (2005) Cotton germplasm status report and crop vulnerability status. Report of Cotton crop germplasm committee January 20, 2005. http://www.ars-grin.gov/npgs/cgc_reports/cottonstatus2005final.pdf
- Anonymous (2006) Multilocation evaluation of cotton genetic resources- Annual Progress Report 2005–2006. In the Annual Group Meeting held at Agricultural Research Station, University of Agricultural Sciences, Dharwad on 7–9 April 2006. Central Institute of Cotton Research, Nagpur, India
- Ansingkar, A.S., Kadke, P.P., Borikar, S.T. and Bhosle, S.S. (2004). Altering *G. hirsutum* cotton at cellular level to impart multiple sucking pest resistance through interspecific hybridization. In: B. M. Khadi, I.S.Katageri, S.S.Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), Proceedings of International symposium on "Strategies for Sustainable Cotton Production- A Global Vision"1, Crop Improvement, 23–25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp. 101–103.
- Babu, V. R., Punith Mohan, Singh, P. and Singh, V. V. (2003) Salt tolerance in the germplasm of diploid cottons. Indian J. Agric. Sci. 73.
- Batade, S.S. (1983) Environmental influence on the magnitude of heterosis in *Gossypium arboreum* L. Indian J. Agric. Sci. 53(8), 627–633.
- Brubaker, C.L., Bourland, F.M. and Wendel, J.F. (1999) The origin and domestication of cotton. In: C.W. Smith and J.T. Cothren (Eds.) Cotton. *origin, history, technology and production*. John Wiley, New York, pp. 3–31.
- Chinchane, B.N., Nandeshwar, S.B., Deshpande, L.A. and Chinchane, V.N. (2004) Agrobacterium mediated transformation and regeneration in *Gossypium arboreum* L. cotton (cv. PA 225). In: B. M. Khadi, I.S.Katageri, S.S. Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), Proceedings of International symposium on "Strategies for Sustainable Cotton Production- A Global Vision" 1, Crop Improvement, 23–25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp. 452–456.
- Deshpande, L.A., Baig, K.S., Kulkarni, V.N., Matishchandra and Kakde, S.S. (2004) Strategies to improve cultivated diploids –Perspective. In: B. M. Khadi, I.S. Katageri, S.S. Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), Proceedings of International symposium on "Strategies for Sustainable Cotton Production- A Global Vision"1, Crop Improvement, 23–25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp. 25–30.
- Deshpande, L.A., Kokate, R.M., Kulkarni, U.G. and Nerkar, Y.S. (1992) Cytomorphological studies in induced tetraploid *G. arboreum* and new interspecific hybrid between 4n *G. arboreum* X *G. hirsutum* . In Proceedings of the First Vasant Rao Naik Memorial National Seminar on Cotton Development held between 5–6 December, 1992. Nagpur, India.
- Deshpande, L. A., Narula, A.M. and Baig, K.S. (2001) Can newly developed quality desi cottons (*G. arboreum*) substitute hirsutums. In: A. M. Narula and Anupam Barik (Eds), Book of Papers, Proceedings of National Seminar on Sustainable Cotton Production to Meet the Requirement of Industry held on 3–4, October. Mumbai, India. pp. 55–74.
- Dessauw, D. and Hau, B. (2006) Inventory and history of the CIRAD cotton (*Gossypium spp.*) germplasm collection. Plant Genetic Resource Newsletter.147, 52–58.
- Doak, C.C. (1934) A new technique in cotton hybridizing suggested changes in existing methods of emasculation and bagging of cotton flowers. J. Hered. 25, 201–204.

- Fryxell, P.A. (1979) *The natural history of the cotton tribes*. Texas A and M University Press College Station, USA, XVII, pp. 245.
- Fryxell, P.A. (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheede*. 2, 108–165.
- Gerstel, D.U. (1953) Chromosome translocations in interspecific hybrids of the genus *Gossypium*. *Evolution*. 7, 234–244.
- Gerstel, D.U. and Sarvella, P.A. (1956) Additional observations on chromosomal translocations in cotton hybrids. *Evolution*. 10, 408–414.
- Gulati, A.N. and Turner, A.J. (1928) A note on the early history of cotton. *Textile Inst.* 20, T1–T9.
- Guo, W. Z., Zhou, B. L., Yang, L. M., Wang, W. and Zhang, T. Z. (2006) Genetic diversity of landraces in *Gossypium arboreum* L. race sinense assessed with simple sequence repeat markers. *J. Integrative Plant Biology*. 48(9), 1008–1017.
- Hammer, K., Cifarelli, S. and Perrino, P. (1986) Collection of land-races of cultivated plants in South Italy. *Genetic Resources and Crop Evolution*. 34(2), 261–273.
- Hutchinson, J.B. (1954) New evidence on the origin of the old world cotton. *Heredity*. 8, 225–241.
- Hutchinson, J.B. (1959) *The application of genetics of cotton improvement*. Cambridge University Press, Cambridge.
- Hutchinson, J.B., Silow, R.A. and Stephens, S.G. (1947) *The evolution of Gossypium*. Oxford University Press London.
- ICAC (1999) Cotton production practices. Technical Information Section, International Cotton Advisory Committee. Washington DC, USA
- ICAC (2005) Cotton production practices. Technical Information Section, International Cotton Advisory Committee. Washington DC, USA
- Khadi, B.M. and Kulkarni, V. N. (1999) Cotton. In: V. L. Chopra, (Eds) *Breeding Field Crops -Theory and Practice*, Oxford IBH Publication, New Delhi, India, pp. 531–575.
- Khadi, B.M. Kulkarni, V.N. Bisankoppa, M.S. and Vamadevaiah, H.M. (2004) Introgression derivatives from *Gossypium arboreum* (A₂) and *G.hirsutum* (AD1) serve as source to identify 'D' genome genes – A study. Book of Abstracts of International Cotton Genome Initiative 2004, Hyderabad, India. 10–13 October, 2004, pp. 85.
- Khadi, B. M., Kulkarni, V. N., Manjula, S. M. and Katageri, I. S. (2003) Utilization of exotic genetic resources in genetic improvement of diploid cultivated cotton and delineation of gene pool concepts. . In: A. Swanepoel(Eds), *Proceedings of the World Cotton Research Conference-3 on "Cotton Production for the New Millennium*, 9–13 March 2003, Cape Town, South Africa, pp. 88–102.
- Kulkarni, V.N. (2002) Hirsutisation of *G. arboreum* cotton and genetic emendation of *G. hirsutum* for sucking pest resistance. Ph. D. Thesis submitted to University of Agricultural Sciences Dharwad, India.
- Kulkarni, V.N., Khadi, B.M., Bisankoppa, M.S., Hussain, G.F.S. and Narayanan, K. (2004) Studies on inter-genomic hybridization between A and B genomes of cotton and useful genetic stocks for biotechnological studies. Book of Abstracts of International Cotton Genome Initiative-2004, Hyderabad, India. 10–13 October 2004. pp. 102–103.
- Kulkarni, V.N., Khadi, B.M., Deshpande, L.A., Sreenivasan, S. and Bisankoppa, M.S. (2003) Hirsutisation of diploid cultivated cotton. In the *Proceedings of World Cotton Research Conference-III, Cotton Production for the New Millennium*, Cape Town, Republic of South Africa, 9–13 March 2003. pp. 85–94.
- Kulkarni, V. N., Khadi, B. M. and Sangam, V. S. (2001) Pre- breeding efforts for low gossypol seeds and high gossypol plant in *G. herbaceum* L. cotton utilizing *G. australe* Mueller. *Current Science*. 82, 434–438.
- Kumar, C.R.A. and Rajmani, A. (1994) Estimation of variability and association in karunganni cotton. *Madras Agric. J.* 81 (5), 288–289.

- Laghetti, G., Hammer, K., Cifarelli, S., Branca, F., Diederchsen, A. and Perrino, P. (2002) Collection of crop genetic resources in Egadi archipelago and southern Sicily. *Plant Genetic Resources Newsletter*. 132, 39–47.
- Lather, B.P.S., Chhabra, R.S., Sangwan, B.S. and Siwach, S.S. (2001) Hybrid cotton in North India – Future prospects. In: A. M. Narula and Anupam Barik (Eds) *Book of Papers, Proceedings of National Seminar on Sustainable Cotton Production to Meet the Requirement of Industry* held on 3–4, October. Mumbai, India. pp. 1–6.
- Le, Q.Q. (1996) Evaluation and use of cotton germplasm. *Proceedings of National Workshop on Plant Genetic Resources in Vietnam*, Hanvi on 28–30 March 1995 pp 145–146.
- Liu, D., Guo, X., Lin, Z., Nie, Y. and Zhang, X. (2006) Genetic diversity of Asian Cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. *Genetic Resources and Crop Evolution*. 53, 1145–1152.
- Manjula, S.M., Khadi, B.M., Katnalli, D.S., Pawar, S.V., Shobha Immadi and Katageri, I.S. (2004) Improvement of *G. herbaceum* through introgression breeding. In: B. M. Khadi, I.S. Katageri, S.S. Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), *Proceedings of International symposium on “Strategies for Sustainable Cotton Production- A Global Vision” 1*, Crop Improvement, 23–25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp. 296–299.
- Mehetre, S.S., Shinde, G.C. and Pardeshi, S.U. (2004) Status and strategies of host plant resistance for biotic stress in cotton. In: B. M. Khadi, I.S. Katageri, S.S. Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), *Proceedings of International symposium on “Strategies for Sustainable Cotton Production- A Global Vision” 1*, Crop Improvement, 23–25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp. 31–47.
- Mehtha, N.P. and Patel, C.J. (1983) Hybrid seed production technique for exploitation of heterosis in cotton. *Indian Farm*. 33(9), 23–31.
- Meshram, L.D. and Wadodkar, M.B. (1992) Male sterility in Asiatic cotton *Gossypium arboreum* L. *Hybrid Cotton News Letter*. 2(2), 3.
- Narula, A.M., Acharya, S., Khadi, B.M. and Kulkarni, V.N. (2001) *Gossypium herbaceum* cotton in India. Directorate of Cotton Development, Department of Agriculture and Co-operation, Mumbai, India.
- Ndungo, V., Demol, J. and Marechal, R. (1988) L'ame'lioration du cotonnier *Gossypium hirsutum* L. par hybridation interspecificue. Publications agricoles No.23. Faculte' des sciences agronomiques. De L'Etat, Gembloux, Belgium.
- Omran, A., Asadollah, A. and Saeid, N. (2007) Intragenomic diversity and geographical adaptability of diploid cotton species revealed by cytogenetic studies. *African Journal of Biotechnology* 6(12), 1387–1392.
- Phillips, L. L. (1961) The cytogenetics of speciation in Asiatic cotton. *Genetics*. 46, 77–83.
- Rajput, J.P., Meshram, L.D., Kalpande, H.V., Golhar, S.R. and Bharad (1998) Heterosis studies in Asiatic cotton. *J. Soils and Crops*. 7(2), 166–170.
- Rana, M. K. and Bhat, K. V. (2002) Genetic diversity analysis in Indian diploid cotton (*Gossypium spp.*) using RAPD markers. *Indian J. Genet.* 62(1), 11–14.
- Rana, M. K. and Bhat, K. V. (2004) Analysis of molecular variance in cotton (*Gossypium spp.*) using RAPD markers. *Indian J. Genet.* 64(1), 85–86.
- Rao, N.G.P., Deshpande, L.A. and Khadi, B.M. (2004) Improvement of Asiatic cottons in India. *International Cotton Genome Initiative ICGI 2004, workshop Souvenir*. 48–53.
- Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., et al. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*. 166, 389–417.
- Sandhu, B.S. (1989) Development of desi cotton and its present status in the Punjab. *J. Cotton Res. and Dev.* 3(2), 123–129.

- Seelanan, T., Schnabel, A. and Wendel, J. F. (1997) Congruence and consensus in the cotton tribe. *Syst. Bot.* 22, 259–290.
- Sheidal, M. and Alishah, O. (1998) Morphometric studies of *Gossypium herbaceum* cultivars in the Iran National Genebank. *Plant Genet African Journal of Biotechnology*. 6(12), 1387–1392.
- Silow, R.A. (1944) The genetics of species development in Old World cottons. *J. Genet.* 46, 62–77.
- Singh, V.V. (1983) Range of variability in *Gossypium herbaceum* germplasm. *Cott. Dev.* 14, 45.
- Singh, V. V. (2007) Principal Scientist cotton, Germplasm curator, Central institute of Cotton Research, Nagpur, India. Personal communication.
- Singh, Phundan (1986) Variability in technological characters in *Gossypium arboreum* L. *Indian J. of Agric. Sci.* 56(4), 234–235.
- Singh, Phundan (1998) Cotton breeding. Kalyani publishers. New Delhi, India.
- Singh, V.V., Baitule, S.J. and Pathak, B. R. (2001) Germplasm strength and its utilization in cotton improvement. In: A. M. Narula and Anupam Barik (Eds) Book of Papers, Proceedings of National Seminar on Sustainable Cotton Production to Meet the Requirement of Industry held on 3–4, October. Mumbai, India. pp. 44–54
- Singh, V.V., Khadi, B.M., Kulkarni, V.N., Mohan, P. and Anjali, K. (2004) Cotton. In B. S. Dhillon, R. K. Tyagi, S. Saxena and A. Agarwal (Eds) *Plant Genetic Resources: Oilseed and Cash Crops*. Narosa publishing house. New Delhi, India. pp. 163–183.
- Singh, T.H., Khandola, H.S. and Nagi, P.S. (1975) Hybrid vigour in intervarietal crosses of desi cotton (*G. arboreum* L.). *J. Res. PAU, Ludhiana, India.* 12(3), 218–223.
- Singh, V. V., Mohan, P., Kulkarni, V. N., Baitule, S. J. and Pathak, B. R. (2003) Explorations within India for collection of cotton species germplasm. *Plant Genetic Resource Newsletter*. 136, 40–46.
- Singh, Phundan and Nandeshwar, S.B. (1983) Variability in *G. arboreum* L race cernum in Garo Hills of India. *Indian J. Agric. Sci.* 53(7), 511–513.
- Singh, Pundan and Narayanan, S.S. (1987) Genetics of locule retentivity in *Gossypium arboreum* Linn. *Indian J. Agric. Sci.* 57(7), 453–455.
- Singh, D.P., Rameshkumar and Lather, B.P.S. (1992) Use of GMS system for economic hybrid seed production in desi cotton. In Proceedings of the First Vasant Rao Naik Memorial National Seminar on Cotton Development held between 5–6 December, 1992. Nagpur, India. 10–17.
- Singh, Phundan and Singh, J. (1984) Variability for some economic characters in the genetic stocks of *Gossypium arboreum* and *G. barbadense* cottons. *Cotton Dev.* 14, 15–16.
- Stanton, M.A., Stewart, J.Mc. D., Percival, A. E. and Wendel, J.F. (1994) Morphological diversity and relationships in the A – genome cottons, *G. arboreum* and *G. herbaceum*. *Crop Sci.* 34, 519–527.
- Wendel, J. F. (1989) New world tetraploid cottons contain old world cytoplasm. *Proc. Natl. Acad. Sci. USA.* 86, 4132–4136.
- Wendel, J. F., Olson, P. D. and Stewart, J. Mc. D. (1989) Genetic diversity, introgression and independent domestication of old world cultivated cottons. *Amer. J. Bot.* 76, 1795–1806.
- Xiang, X. T. (1988) Study and utilization on Asiatic cotton (*G. arboreum*) in China. *Sci Agr. Sin.* 21, 94.

Gossypium DNA Markers: Types, Numbers, and Uses

Mehboob-ur-Rahman, Yusuf Zafar, and Andrew H. Paterson

Abstract Numerous DNA marker systems, each with their own advantages and disadvantages, have been described for cotton. Microsatellite markers (also known as simple sequence repeats [SSRs]) are robust and, coupled with their co-dominant expression, make them the current marker of choice for phylogenetic/evolutionary studies, development of genome maps and the identification of DNA markers linked to desirable traits for use in marker-assisted selection (MAS) in cotton. Despite the advances made with such markers, there remains a need for new genomic tools, such as single nucleotide polymorphisms (SNPs), to identify further polymorphism in cotton. The allo-ploidy of the leading cultivated cotton species, together with its widespread dispersal and the selection pressure applied during domestication and scientific improvement, has resulted in a very narrow genetic base. There is thus a need to introgress novel traits from feral or wild relatives, requiring suitable markers for use in MAS. In the future more attention will be devoted to producing consensus genome maps, which will be important if the aim of ‘breeding by design’ is to be realized in cotton improvement.

1 Introduction

Cotton is the world’s most important natural textile fiber (Rahman et al. 2008a). Future gains in fiber yield and quality are expected to be accelerated by employing modern genomic tools to discover DNA polymorphisms and utilize them in cotton breeding in a process called marker assisted selection (MAS). The marker technology has enabled breeders to utilize DNA markers for selecting desirable plants without exposing them to a certain set of environments (Helentjaris et al. 1986), paving the way for adoption of this technology

Mehboob-ur-Rahman (✉)

Plant Genomics & Mol Breeding Labs. National Institute for Biotechnology & Genetic Engineering (NIBGE) Faisalabad Pakistan
e-mail: mehboob_pbd@yahoo.com

by different entrepreneurs including the public sector (Welsh and McClelland 1990; Vos et al. 1995; Struss and Plieske 1998).

The cotton genome is large and complex with 26 chromosomes, necessitating the identification of large number of DNA markers for gene tagging and genome mapping; for localizing qualitative and quantitative traits (Park et al. 2005), genetic diversity studies, varietal protection and pedigree analysis; and for introgressing novel traits from unadapted germplasm (Paterson et al. 1991; Qureshi et al. 2004).

The range of DNA markers that can easily be used is quite extensive. Techniques applied to study plant germplasm variations include identifying polymorphisms in the actual DNA sequence, the use of DNA hybridization methods to identify restriction fragment length polymorphisms (RFLPs), or the use of PCR-based technology to find polymorphism using random amplified polymorphic DNA (RAPD), simple sequence repeat (SSR), or amplified fragment length polymorphism (AFLP) (Hamada et al. 1982, Williams et al. 1990; Reinisch et al. 1994; Vos et al. 1995). These DNA markers have been developed and used in cotton for fingerprinting (Rahman et al. 2002b, 2008b), linkage map construction (Reinisch et al. 1994; Ulloa et al. 2002; Zhang et al. 2002; Lacape et al. 2003; Mei et al. 2004; Rong et al. 2004), gene mapping (Shapley et al. 1998; Ulloa and Meredith 2000), and genetic diversity studies (Rahman et al. 2002b, 2008b). DNA marker identification in cultivated cotton is constrained by its narrow genetic base (Abdalla et al. 2001; Rahman et al. 2002b; Rahman et al. 2005a). However, a large number of markers is needed to construct a saturated genetic map (Han et al. 2004). A high-density genetic map can assist breeding procedures by directly tagging the genes conferring traits of interest. To date, several genetic maps of tetraploid cotton have been constructed using diverse molecular markers and mapping populations (Reinisch et al. 1994; Ulloa et al. 2002; Lacape et al. 2003; Rong et al. 2004).

2 Restriction Fragment Length Polymorphism (RFLP)

2.1 Basis of the Method

In RFLP analysis, organisms are differentiated by analysis of fragments generated after cleaving their genomic DNA with restriction enzymes. The polymorphisms are the result of variations in restriction sites or intervening sequences among different organisms yielding DNA fragments of different length (Botstein et al. 1980). The main steps involved in RFLP analysis consist of isolation of DNA, digestion of DNA with restriction enzymes, separation of restriction fragments by agarose gel electrophoresis, transfer of the restriction fragments to a filter by Southern blotting, detection of the restriction fragments with probes, and scoring of segregation by direct observation of autoradiograms. RFLP markers are reliable; their gene dosage is usually co-dominant,

and is simply inherited (Paterson and Zhao 1995; Kohel et al. 2001; Mei et al. 2004) (Table 2). RFLPs have been extensively used to study genetic diversity, population genetics, evolutionary history, and genome mapping in many crop species (Dudley et al. 1992; Giese et al. 1993; Paterson et al. 1993; Schon et al. 1993; Shappley et al. 1998).

Several research groups have utilized the RFLP technique in cotton (Meredith 1992; Reinisch et al. 1994; Jiang et al. 2000; Mei et al. 2004; Ulloa et al. 2005). It has been used for molecular differentiation at the species level (Reinisch et al. 1994; Mei et al. 2004) and at the intraspecific level (Becelaere et al. 2005). An early evaluation of RFLP markers was based on studying heterosis and varietal origins in Upland cotton, *Gossypium hirsutum* L., employing 75 probe/enzyme combinations on 68 individual lines of diverse origin and resulting in 179 polymorphic fragments (Meredith 1992).

2.2 Application to Phylogenetic and Diversity Studies

Evolutionary inferences have been drawn by comparing allotetraploids and their diploid progenitors using the RFLP procedure. It was reported that recombination within each of the A and D genomes was increased in tetraploid cotton, compared with the respective diploid cottons while retaining the same physical length of A and D genomes in both the tetraploid and diploid (Brubaker et al. 1999).

RFLP analysis was applied to understand the extent of genetic diversity among *G. hirsutum* cultivars released in the USA between 1970 and 1990. A greater level of genetic diversity was estimated in cotton than in the other crops (Bowman et al. 1996). However, it was pointed out that the level of genetic diversity of cotton might be overestimated (Van Esbroeck et al. 1999). Low genetic diversity was found in other studies using isozymes (Wendel et al. 1992) and DNA markers (Brubaker and Wendel 1994). RFLP-based genetic similarity estimates are more accurate estimates of true genetic resemblance among cotton cultivars than inferred relatedness from pedigree analysis (Bacelaere et al. 2005).

2.3 Application to Linkage Mapping

The first detailed cotton RFLP map was reported using a population derived from an interspecific cross between *G. hirsutum* race Palmeri and *G. barbadense* acc. K-101 (Table 1). A total of 705 RFLP loci were sorted into 41 linkage groups, covering 4,675 cM of the cotton genome, and 14 linkage groups were associated with numbered chromosomes using a series of monosomic interspecific substitution stocks. RFLP markers were slightly more abundant in the D subgenome than the A subgenome (Reinisch et al. 1994). It was also

Table 1 Reported molecular marker based linkage maps in cotton for (a) Intrasppecific (b) Interspecific

Marker	Parentage	G C ⁺ (cM)	No. of LG ^{a,b}	References
A- Intrasppecific population (<i>G. hirsutum</i> x <i>G. hirsutum</i>)				
RFLP	HS46 x MARCABUCAG8US-1-88	43	5	Shappley et al 1994
RFLP	HS46 x MARCABUCAG8US-1-88	865	31	Shappley et al 1996
RFLP	MD5678ne x Prema	700	17	Shappley et al 1998
RFLP	HQ95-6 x MD51ne	1503	47	Ulloa and Meredith 2000
SSR, AFLP and morphological	Yumian-1 x T-586	525	20	Ulloa et al 2002
EST-SSR	(TM-1 x Hai 7124)TM-1	–	6	Zhang et al 2005
SSR	7225 x TM-1	1024.4	–	Han et al 2006
RAPD and SSR	FH-63IS x PBD-883	274.1	4	Shen et al 2007
B- Interspecific Population (<i>G. hirsutum</i> x <i>G. barbadense</i>)				
RAPD and RFLP	<i>G. barbadense</i> K-101 x <i>G. hirsutum</i> Palmeri	6663	51	Asif 2007
RFLP	<i>G. hirsutum</i> Empire B2b6 x <i>G. barbadense</i> Pima S-7	4675	41	Reinish et al 1994
RFLP	<i>G. hirsutum</i> Tamcot 2111 x <i>G. barbadense</i> Pima S-7 BC ₃ F ₂	–	–	Khan et al 1998
RAPD, AFLP and morphological	<i>G. hirsutum</i> TM-1 x <i>G. barbadense</i> Pima 3–79	521.5	11	Wright et al 1999
RFLP and isozyme	<i>G. hirsutum</i> x <i>G. barbadense</i>	856	18	Saranga et al 2001
RFLP and RAPD	<i>G. hirsutum</i> TM-1 x <i>G. barbadense</i> 3–79	1486	17	Chee et al 2005a,b
RFLP	<i>G. hirsutum</i> CAMD-E x <i>G. barbadense</i> See Island Seaberry	4766	50	Darye et al. 2005
SSR and RAPD	<i>G. hirsutum</i> x <i>G. barbadense</i>	3664	26	Altai et al 1997
RAPD and SSR	<i>G. hirsutum</i> x <i>G. barbadense</i>	3312.2	42	Brubakar et al 1999
RAPD and SSR	NM240616 x <i>G. hirsutum</i> TM-1	331.5	43	Kohel et al 2001
		1502.6	16	Jiang et al 2000
				Guo et al 2002
				Zhang et al 2002
				Ulloa et al 2002

Table 1 (continued)

Marker	Parentage	G C* (cM)	No. of LG**	References
SSR	<i>G. hirsutum</i> x <i>G. barbadense</i> BC1	2126.3	46	Frelichowski et al 2006
RFLP, AFLP and SSR	<i>G. hirsutum</i> x <i>G. barbadense</i>	5500	37	Lacape et al 2003
RFLP, AFLP and SSR	<i>G. hirsutum</i> Acala 44 x <i>G. barbadense</i> Pima S-7	3287	42	Mei et al 2004
SSR	<i>G. hirsutum</i> TM-1 x <i>G. barbadense</i> Hai 7124	5519	–	Nguyen et al 2004
SSR, CSR and EST-SSR	<i>G. hirsutum</i> TM-1 x <i>G. barbadense</i> Pima 3–79	1277	11	Wang et al. 2008
SSR, SRAP and RAPD	<i>G. hirsutum</i> Handan 208 x <i>G. barbadense</i> Pima 90	5141.8	41	Park et al 2005
SSR	<i>G. hirsutum</i> TM-1 x <i>G. barbadense</i> Hai 7124	4331.2	34	Lin et al 2005
SSR, SRAP, RAPD and REMAP	<i>G. hirsutum</i> Handan 208 x <i>G. barbadense</i> Pima 90	5472.3	26	Song et al 2005
SSR	<i>G. anomalum</i> -7235 x <i>G. hirsutum</i> TM-1	1024.4	–	He et al 2007
SSR	<i>G. hirsutum</i> FH-1000 x <i>G. barbadense</i> PGMB-36	–	–	Shen et al 2007

*GC: Genome coverage; **LG = Linkage group

concluded that the cotton genome contains about 400-kb DNA per cM on average. The cotton map afforded new opportunities to study chromosome evolution, and to exploit *Gossypium* genetic resources.

Small differences were found in the linear order of the genetic map (RFLP markers made 27 linkage groups) constructed on an F_2 population comprised of 271 individuals derived from a cross of *G. hirsutum* cv. "CAMD-E" \times *G. barbadense* cv. "Sea Island Seaberry" compared to the published map by Reinisch and Co-workers (Jiang et al. 1998). Similarly, 355 DNA markers (216 RFLP and 139 RAPD) linked with fiber quality properties, assembled into 50 linkage groups covering around 4766 cM, were segregating in a population comprising 171 F_2 individuals from a cross between *G. hirsutum* acc. TM-1 and *G. barbadense* acc. 3-79 (Kohel et al. 2001). In many efforts, some polymorphic RFLPs could not be assembled into linkage groups (Shapple et al. 1998) because of the fact that segregation distortion and linkage drag may affect locus behavior among different interspecific hybrid populations (Reinisch et al. 1994; Mei et al. 2004).

In another study, an interspecific backcross population comprising 24 BC_3F_1 (backcross- $3F_1$) plants was surveyed with 262 RFLP markers selected for even coverage of the cotton genome. A subset of 127 RFLP markers was found sufficient to monitor all the introgressed regions (Chee et al. 2000). Similarly, an interspecific backcross population (75 BC_1) of cultivated tetraploid cotton was evaluated with 1014 markers assembled into 37 linkage groups. The overall congruency in the locus orders and distances of common SSR and RFLP loci in these maps allowed for an estimation of 5500 cM as a minimum consensus length (Lacape et al. 2003).

The first RFLP genetic linkage map based on an intraspecific population derived from a cross of two *G. hirsutum* genotypes (Table 1), was comprised of five linkage groups containing 12 loci (Shapple et al. 1996). Two years later, a more-detailed linkage map was published on this population, with a total of 120 loci grouped into 31 linkage groups covering 865 cM (18.6%) (Shapple et al. 1998). In another study, a genetic linkage 'joinmap' was constructed on four different mapping populations of *G. hirsutum*. Out of these, two populations derived from crosses HS46 \times MAR (Shapple et al. 1998) and MD5678ne \times Premma (Ulloa and Meredith 2000) have previously been reported. On the two new populations (Table 1) maps were developed comprising 83 loci mapped to 24 linkage groups with an average distance between the markers of 10 cM, covering 830.1 cM. The second genetic map encompassed 56 loci mapped to 16 linkage groups with an average distance between markers of 9.3 cM, covering 520.4 cM or approximately 11% of the cotton genome. The 'joinmap' from integrating maps of the four populations comprised of 284 loci mapped to 47 linkage groups with the average distance between markers of 5.3 cM, covering 1,502.6 cM or approximately 31% of the total cotton genome. The joinmap provided further knowledge of comparative chromosome arrangement, parental relationships, gene order, and increased the potential to map genes for

cotton improvement (Ulloa et al. 2002). Chromosome identities were assigned to 15 linkage groups.

2.4 Application to Gene Tagging and QTL Mapping

RFLP mapping has been used in many studies to dissect the genetic control of complex traits in cotton. A first detailed study has elucidated the role of the D-subgenome (from a progenitor that does not produce spinnable fibers) in imparting fiber and yield-influencing quantitative trait loci (QTLs) to the allotetraploid cultivated cotton species. A total of 14 QTLs were identified using an F_2 population (Table 1), most located on the D-subgenome. This study also shed light on differential response to selection of the fiber and productivity associated QTLs in the cultivated allotetraploid and A-genome diploid species (Jiang et al. 1998).

An RFLP genetic linkage map on $F_{2:3}$ progeny derived from a cross of *G. hirsutum* (Table 1) was constructed, comprising 81 loci mapped to 17 linkage groups covering 700.7 cM of the cotton genome with an average distance of 8.7 cM between the markers. Twenty six QTLs representing different agronomic and fiber quality traits were detected on nine linkage groups. The QTL positions on the linkage groups suggested that genes conferring fiber quality might cluster on the same cotton chromosome(s) (Ulloa and Meredith 2000).

In multiple investigations on a backcross population (BC_3F_2) derived from a cross of *G. hirsutum* cv. Tamcot 2111 and *G. barbadense* cv. Pima S6 followed by backcrossing of the F_1 plants to the *G. hirsutum* parent for three cycles were field-tested independently for different fiber quality parameters. These progenies were genotyped with RFLP markers. A total of 22 non-overlapping QTLs distributed over 15 different chromosomes for fiber elongation (Chee et al. 2005a); 32 and 9 QTLs for fiber fineness and micronaire, respectively (Draye et al. 2005), and 28, 9, and 8 QTLs for fiber length, length uniformity and short fiber content (Chee et al. 2005b), respectively, were identified. Correlations among QTLs influencing fiber length, length uniformity, and short fiber content were also reported. These DNA markers associated with QTLs coming from *G. barbadense* may help in breeding for high quality lint (Chee et al. 2005b).

Water stress is one of the major factors limiting the cotton productivity. Saranga and Co-workers dissected the genetics of different physiological and productivity traits under limited water regime. Out of 161 QTLs, only 33 QTLs significantly influenced the plant productivity (11 QTLs), and physiological (five QTLs) and fiber quality (17 QTLs) traits (Saranga et al. 2001). QTLs conferring complex traits such as fiber quality were prone to fluctuating environments such as growing season and variable water regimes (well-watered versus water-limited). A total of 6, 7, 9, 21, 25 and 11 QTLs determining the fiber length, length uniformity, elongation, strength, fineness, and color, respectively, were identified in one or more growth conditions (Paterson et al. 2003).

Four QTLs were mapped affecting density of leaf and stem trichomes on an F_2 population derived from a cross between a leading *G. barbadense* cv. Pima S-7 and *G. hirsutum* cv. Empire B2b6. A QTL on chromosome 6 for dense leaf pubescence was inferred to be the t_1 locus; a second QTL on chromosome 25 (homoeologous to chromosome 6) fit the description of the t_2 locus and the two additional QTLs explained significant phenotypic variation in leaf pubescence (represented as t_3 , t_4 , t_5 loci). It was suggested that different genes control the trichome density on the stem and leaf (Wright et al. 1999).

Findings from the 'joinmap' described above (section 1.3) were integrated with QTL identification for multiple fiber and yield traits, concluding that 63 QTLs were present on five different chromosomes of the A-subgenome (chromosomes 3, 7, 9, 10 and 12) and 29 QTLs on three different D-subgenome (chromosomes 14 Lo, 20, and the long arm of 26). Most of the QTLs were clustered on just two chromosomes (Ulloa et al. 2005).

A total of 226 RFLPs were employed to identify markers associated with resistance to *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) using four different populations developed from crosses between four resistant *G. hirsutum* parents and a single *G. barbadense* parent, "Pima S-7," highly susceptible. Four maps containing 162, 224, 253, and 184 loci (one for each population) were assembled into 48, 45, 49, and 48 linkage groups, respectively. Seven QTLs for *Xcm* were detected among the four populations (Wright et al. 1998).

2.5 Summary - Strengths and Weaknesses of the Method

The work carried out to date gives an indication that RFLP markers can be useful for discovering polymorphisms in the cotton genome (Wang et al. 1995; Mei et al. 2004). Indeed, RFLPs are the backbone of several cotton reference maps, have improved knowledge of genome organization, and were the basis of most early QTL mapping studies. However, despite these positive results, RFLPs have not been widely used in recent years because of low polymorphism in cotton compared to the other plant taxa (Brubaker and Wendel, 2000). Also, the technique requires large quantities of DNA, and is costly, time-consuming and difficult to automate (Beckmann and Soller 1983; Williams et al. 1990).

3 Random Amplified Polymorphic DNA (RAPD) and other Arbitrary-Sequence PCR Methods

3.1 Basis of the Method

To counteract the disadvantages of RFLP, a PCR based marker system was developed which utilizes the arbitrarily primed regions to scan polymorphisms, avoiding the need for prior knowledge of nucleotide sequences or cloned and

characterized hybridization probes (Bassam et al. 1995). The PCR-based marker assays also require smaller quantities of DNA than RFLP (Williams et al. 1990). The PCR-based assays yield characteristic amplification signatures (fingerprints) from all types of genomic DNA using at least one short oligonucleotide of arbitrary or semi-arbitrary sequence as a primer. Usually, the length of the primer varied between 10-mers (Williams et al. 1990) to 20-mers (Welsh and McClelland 1990); attempts to use 5 or 6-mers yielded irreproducible results. RAPD (Williams et al. 1990), arbitrarily primed PCR (AP-PCR; Welsh and McClelland 1990) and DNA amplification fingerprinting (DAF; Caetano-Anollés et al. 1991) have been used to generate PCR-based fingerprints. The basic principles of RAPD, AP-PCR and DAF are the same; however, the difference stands with the random primer length and gel electrophoresis. For example, in RAPD analysis 10-mer primers are used while DAF uses 7–8 nucleotides long (but in some cases as short as 5 nucleotides) and AP-PCR is based on 20-base primers. Another important factor which differentiates these techniques is the primer and template concentrations. DAF requires primer/template ratio greater than 5 (at least 5:1), and RAPD requires a ratio less than or equal to 2. AP-PCR falls in between. The DAF assay yields a larger number of amplification products than RAPD and AP-PCR, which are separated by polyacrylamide gel electrophoresis (PAGE). Silver staining is used for DNA detection in DAF, while autoradiography is used for AP-PCR. In the RAPD procedure, agarose gel electrophoresis is used followed by detection of DNA fingerprints with ethidium bromide.

The advantages of the RAPD method are: 1) A single primer of random sequence can be used for all types of organisms. 2) One avoids the need for isolation of cloned DNA probes, radioactivity and Southern transfer. 3) The genotyping procedure can be automated (Table 2). The RAPD procedure has remained the most popular in studying polymorphisms in different plant taxa (Klein-Lankhorst et al. 1991; Rahman and Zafar 2001; Mukhtar et al. 2002; Rahman et al. 2004; Asif et al. 2005, 2006; Tabbasam et al. 2006). Sequence variation in the genome alters the primer binding sites which are the main cause of polymorphism, making RAPDs dominant markers. They are most often used as markers for diversity, genome mapping and polygenic studies in cotton (Rahman et al. 2002b; Zhang et al. 2002; He et al. 2007). Like other molecular markers, RAPDs were extensively used to tag genes, to fingerprint genomes, and to produce genomic maps.

3.2 Application to Phylogenetic and Diversity Studies

Several workers have investigated phyletic relationships in the genus *Gossypium* using analysis of morphology (Fryxell 1971; Valicek 1978), floral flavonoids (Parks et al. 1975), seed protein electrophoretic patterns (Cherry et al. 1970; Johnson and Thein 1970), interspecific meiotic pairing behavior (Phillips 1966),

Table 2 Molecular markers and their advantages and disadvantages

Molecular marker	Advantages	Disadvantages
RFLP	<ul style="list-style-type: none"> • Co-dominant • Provides many conserved-sequence markers used for comparative analysis within and among genomes • No prior sequence information needed, however sequence of probes often provides information needed for PCR-based assay • Markers are well-distributed throughout the genome • Often obtained multiple bands per experiment • Robust assay 	<ul style="list-style-type: none"> • Requirement of high quality DNA • Large amount of DNA required • Slow • Automation difficult • Standardization difficult • Statistical analysis difficult • Low level of polymorphism • Frequent use of radioisotopes in detection • Technically demanding and laborious
RAPD	<ul style="list-style-type: none"> • Less quantity (5-50 ng per reaction) of template DNA • No prior sequence information • Cost effective and easy to use • Generation of multiple bands per reaction • Can be automated • No use of radioactive isotopes 	<ul style="list-style-type: none"> • Purified and high molecular weight DNA required • Contamination of DNA • Needs highly standardized protocol • Interpretation of band profiles in terms of loci and alleles is not possible • Dominant expression • Low reproducibility • Similar sized fragments may not be homolo
SCARs	<ul style="list-style-type: none"> • Simpler patterns than RAPDs (locus-specific) • Robust assay • Co-dominant markers 	<ul style="list-style-type: none"> • Sequence information needed • Require effort and expense in designing specific primers
AFLP	<ul style="list-style-type: none"> • Large number of amplicons (40-50) are generated • No prior sequence information or probe • Can be automated 	<ul style="list-style-type: none"> • High initial cost • Dominant markers • Often cluster at the centromeres and telomeres. • Complex and laborious marker system
SSRs	<ul style="list-style-type: none"> • Require small quantity of genomic DNA • Highly polymorphic • Widely distributed throughout the genome • Easy interpretation in genotyping • Automation is easy, multiplexing is possible • Highly reproducible 	<ul style="list-style-type: none"> • Initial cost is high • Discovery procedure is quite complex
ISSRs	<ul style="list-style-type: none"> • No prior sequence information • Variation within unique regions of the genome may be found at several loci simultaneously 	<ul style="list-style-type: none"> • Dominant markers • Detection system is complex

Table 2 (continued)

Molecular marker	Advantages	Disadvantages
	<ul style="list-style-type: none"> • Tend to identify significant levels of variation • Microsatellite sequence-specific • Useful for DNA profiling for closely related species 	
SNPs	<ul style="list-style-type: none"> • Detection of all types of sequence change • Highly abundant • Co-dominant • Easily automated 	<ul style="list-style-type: none"> • Sequence information is highly needed • High initial cost

allozymes (Wendel et al. 1992), RFLP (Wendel and Brubaker 1993) and DNA sequences of chloroplast, nuclear ribosomal DNA and low-copy nuclear genes (Wendel and Albert 1992; Cronn et al. 1996, 2002; Shaheen et al. 2006). Some genetic differences may not be visualized by the conventional techniques due to the paucity of loci that can be resolved. The RAPD procedure was applied to understand phylogenetic relationships among 31 *Gossypium* species, three subspecies and one interspecific hybrid. Genetic similarity among the species was in the range of 78.8% to 92.30 % with 45 random decamer primers. The interspecific genetic relationship of several species was related to their center of origin (Khan et al. 2000).

In Pakistan, epidemics of cotton leaf curl disease (CLCuD), a disease of viral origin, was the compelling factor to devise new strategies in cotton breeding programs. Genetic similarity among the elite cotton cultivars released prior to CLCuD epidemics was in the range of 81.5% to 93.41%, which was alarming (Iqbal et al. 1997). Later, 20 resistant cotton cultivars were developed by crossing the exotic resistant germplasm (LRA-5166, CP-15/2 and Cedix) with adapted varieties, highly susceptible to the disease. Genetic relatedness among these cultivars was in the range of 81.45% to 94.90%. It has been demonstrated that most of the cultivars possess narrow genetic backgrounds which is an area of concern for many cotton breeders (Bowman 1999; Rahman et al. 2002b; Rahman et al. 2006b). With RAPD analysis, closely related varieties could be distinguished (Multani and Lyon 1995). Moreover, the genetic distance obtained with RAPD markers is comparable with the taxonomic distances measured from morphological features. RAPD analysis is also instrumental in clustering near homozygous cotton lines derived from interspecific hybridization between *G. hirsutum* and *G. barbadense* into two clusters with one resembling *G. hirsutum* and the other *G. barbadense* (Tatineni et al. 1996).

As in *G. hirsutum*, genetic diversity is limited in *G. arboreum* genetic stocks as revealed by RAPD analysis conducted on 30 genotypes. The average genetic relatedness among all the genotypes was 80.46%. None of the primers surveyed

could differentiate all genotypes (Rahman et al. 2008b). These findings are in agreement with earlier studies on diploid cultivars (Rana and Bhat 2004). Most of the genotypes developed from one cotton research institute clustered together. Such commonalities have been found in cotton (Iqbal et al. 1997, 2001; Rahman et al. 2002b). Two genotypes originating from two different countries grouped in one cluster. It is likely that these were developed from a common population (Bligh et al. 1999; Pillay and Myers 1999).

3.3 Application to Linkage and QTL Mapping

3.3.1 Fiber Traits

Introgression from *G. barbadense* into *G. hirsutum* led to a dramatic genetic gain in fiber quality and other important traits, based on two mapping populations (TM1 x NM24016 and NM24016 x 3-79) where 50 RAPDs and SSRs anchored to individual cotton chromosomes were mapped. QTLs in both populations were identified for fiber strength, fiber length, and fiber fineness. The QTLs identified could be used for introgression of novel QTLs into cultivated cotton (Cantrell et al. 1999).

The RAPD analysis was performed on 70 F₂ individuals of a cross between *G. hirsutum* acc. TM-1 and *G. barbadense* acc. 3-79 using approximately 220 ten-mer primers. Thirty four percent of the loci were polymorphic between the parents (Lazo et al. 1994). In another study, 26% of the loci were found polymorphic between *G. hirsutum* var FH-1000 and *G. barbadense* var PGMB-36 (Mumtaz 2007). Most of the markers were parental genome specific, and identified few anchors to known chromosomes (Lazo et al. 1994; Yu et al. 1996). RAPD markers co-segregating with fiber strength QTLs were identified (Yu et al. 1996). A detailed molecular map based on 219 RAPD, RFLP and SSR markers was constructed, and these loci were assembled into 40 linkage groups covering about 3855 cM. About one half of the linkage groups were assigned to their respective genomic origin (A or D) by the use of diploid and aneuploid cottons.

Two QTLs for bundle fiber strength, three QTLs for fiber length, and five QTLs for fiber fineness in different linkage groups explaining about 35 to 50% of the total genetic variance were reported. A total of 355 RFLPs and RAPDs grouped into 50 linkage groups covering 4766 cM of the segregating population developed from a cross between *G. hirsutum* acc. TM-1 and *G. barbadense* acc. 3-79 (Kohel et al. 2001). The size of the RAPD fragments ranged from 250 to 3500 bp, and most were genetically dominant in expression (Kohel et al. 2001).

In another linkage map, 43 linkage groups covering 3314.5 cM using 58 haploid and doubled haploid individuals originating from a cross of *G. hirsutum* accession TM-1 and *G. barbadense* cv Hai 7124 were reported (Table 1; Zhang et al. 2002). The linkage groups were associated with chromosomes of the allotetraploid genome exploiting the monosomic and telodisomic genetic

stocks. Some groups unassociated with chromosomes were connected to corresponding A or D subgenomes. It was revealed that chromosomes 5 and 18 might be a new pair of homoeologous chromosomes of the allotetraploid cotton genome (Zhang et al. 2002).

G. anomalum, another important species characterized for desirable fiber quality traits, along with *G. hirsutum* accession TM-1 and their F₂ and F₃ population were screened with RAPDs and SSRs resulted into the identification of six RAPDs and three SSRs linked with two QTLs for fiber strength which explained 30% of the phenotypic variance (Zhang et al. 2003).

3.3.2 Fertility Restoration

Like RFLPs, RAPDs have been used to identify diagnostic DNA markers in cotton (Rahman et al. 2005a). A RAPD marker at a genetic distance of 6 cM from the male-fertility gene of a male-sterile cytoplasm was discovered in *G. hirsutum* (Lan et al. 1999). Five RAPD markers (UBC147₁₄₀₀, UBC607₅₀₀, UBC679₇₀₀, UBC659₁₅₀₀ and UBC169₈₀₀) were identified using bulked segregant analysis (BSA) closely linked to the two independent dominant restorer genes, *Rf*₁ from the D₂ restorer line transferred from *G. harknessii* Brandegee (D₂ genome) and *Rf*₂ from the D₂ restorer line (Zhang and Stewart 2004). These RAPD markers were converted into STS markers in *G. hirsutum* (Feng et al. 2005). Two RAPD markers, NAU/RAPD/Rf13₁₄₈₀ and NAU/RAPD/Rf15₇₁₀ and three SSR markers closely linked to the *Rf*₁ gene were reported (Liu et al. 2003). A RAPD marker, UBC188₅₀₀ exhibiting a recombination frequency of 5.26% with gametophytic restorer *Rf*₂ was identified which agreed with the previous estimate from a testcross, (D8R x T586) x H1330 (Zhang et al. 2005). These restorer-specific markers would be useful in marker-based selection for developing restorer parental lines and constructing a high-resolution linkage map containing *Rf*₁ and *Rf*₂ (Liu et al. 2003; Zhang and Stewart 2004; Feng et al. 2005; Zhang et al. 2005). The molecular markers found would be helpful in the development of the elite restoring lines in cotton by MAS.

3.3.3 DNA Markers for Physico-Morphological Traits

Many efforts were made to find DNA markers associated with fiber properties, however, limited studies were reported to exploit host plant resistance mechanisms (HPR) against the insect pests and diseases in cotton, though its importance has been realized since the third century BC. Resistance against jassids in *G. hirsutum* genotypes was due to the presence of long dense hair (Knight and Sadd 1954; Butter and Vir 1991; Ali et al. 2004). The pilose hairiness trait also confers resistance to boll weevil, pink bollworm and plant bugs and is neutral for mites (Parnell et al. 1949; Ahmad et al. 1987). Similarly, nectarilessness is another trait in the cotton defense umbrella, which minimizes the attack of bollworms and whitefly (Bhatnagar and Sharma

1991). A number of insect resistance genes have been genetically tagged and mapped using molecular markers (Zhang and Yu 1999). To find DNA markers linked to the velvet hairiness and nectariless trait, BSA with RAPD was used. The average number of RAPDs were 4.98 and 6.47 for velvet hairiness and nectariless traits, respectively (Rahman et al. 2003; Ali 2004). The DNA markers associated with leaf hairiness were OPC-08₇₀₀, OPD-19₆₄₀, OPF-11₆₃₀, OPG-06₉₈₀, OPG-17₅₀₀, OPH-13₁₁₀₀, OPN-14₈₉₀, OPO-11₉₂₀ and OPO-14₁₂₀₀. A preliminary genetic linkage map for hairiness showed that the four RAPD markers OPC-08₇₀₀, OPF-11₆₃₀, OPH-13₁₁₀₀ and OPO-14₁₂₀₀ cosegregated within the locus and exhibited close linkage to the hairiness locus. Three RAPD markers OPG-17₅₀₀, OPO-11₉₂₀ and OPD-19₆₄₀ were present at 1.8 cM on the either side of the locus. One RAPD marker OPN-14₈₉₀ was present at 6.2 cM from the locus (Rahman et al. 2002a; Rahman et al. 2003; Ali, 2004).

Nectariless, another trait of the cotton defense umbrella, is conferred by two dominant genes. RAPD analysis in combination with BSA was applied and one DNA marker OPB-12₁₄₀₀ was identified in repulsion phase. The marker would be helpful to select homozygous F₂ nectariless plants without raising F₃ progeny (Rahman 2002; Rahman et al. 2003). Similarly, after surveying 420 RAPD primers, two DNA markers OPL-09₉₆₀ and OPZ-11₉₅₀ were found associated with the red leaf color trait in *G. hirsutum*. The RAPD marker OPL-09₉₆₀ was linked to the red leaf locus at a distance of 5.0 cM (Rahman 2002; Ali 2004).

3.3.4 Disease Resistance

Cotton cultivation in Pakistan and Northwestern India is severely affected by CLCuD. Three genes are involved in conferring resistance to CLCuD, two for resistance ($R_{1CLCuDhir}$, and $R_{2CLCuDhir}$) and a suppressor of resistance ($S_{CLCuDhir}$) in *G. hirsutum* (Rahman et al. 2005b). The RAPD technique was applied to find DNA markers for CLCuD resistance on a population developed by making a cross between *G. hirsutum* var S-12 (highly susceptible) and *G. hirsutum* var LRA-5166 (resistant). A total of 520 random decamer primers were surveyed using BSA. None of the detected polymorphic DNA fragments were linked to the CLCuD resistance. The analysis was continued on the parents of the population. Thirteen percent of the total amplified DNA fragments were found to be polymorphic among the parents. One DNA marker in trans phase (recombination frequency of 14%) and three markers in coupling phase OPO-19₄₆₀, OPQ-14₃₂₅ and OPY-2₁₀₈₀ (recombination frequency of 0–5%) associated with the disease resistance were identified (Rahman et al. 2006a). In another study, three RFLP markers associated with the disease were identified using an F_{2:3} population derived from a cross between *G. barbadense* (susceptible parent) and *G. hirsutum* (resistant parent) (Aslam et al. 1999).

3.4 RAPD-Derived SCAR Markers

Reproducibility of RAPD markers can be improved by sequencing the termini of diagnostic DNA fragments and designing longer primers for specific amplification of markers (Table 2). In this process, polymorphic fragments amplified by the RAPD are cloned and sequenced to design primers of 20-24 bp length. This strategy converts the dominant RAPD marker into SCAR markers which are co-dominant in expression.

To develop SCAR markers for the detection of UBC431₁₉₂₀ and UBC757₁₃₆₅, the polymorphic DNA fragment with 1920- and 1365-bp size were amplified from *G. anomalum*-7235, and were cloned separately into a vector. Only SCAR431₁₉₂₀ detected polymorphism between TM-1 and 7235. SCAR431₁₉₂₀ was explored to determine its stability in F₂ (7235 x TM-1) with UBC431₁₉₂₀ marker as a control and to screen the major fiber strength QTL of (7235 x Simian-3) BC₁F₄ population for transferring good fiber quality. The result showed that the specific SCAR431₁₉₂₀ marker could be applied to large-scale screening for the presence or absence of the major fiber strength QTL in breeding populations (Guo et al. 2003).

Two RAPD markers, UBC111₃₀₀₀ and UBC188₅₀₀, linked with a fertility restorer gene *Rf*₂ were identified in coupling phase using BSA, and RAPD markers UBC169₇₀₀ and UBC659₁₅₀₀, associated with *Rf*₁ were converted into reliable and genome specific STS markers on the basis of their sequence information. These markers are restorer-specific and could be useful in MAS (Zhang and Stewart 2001; 2004).

3.5 Summary – Strengths and Weaknesses of the Method

Arbitrary PCR-based marker systems are fast, cost effective, easy to use and universal as revealed by many applications (Caetano-Anollés 1996; Rafalski and Tingey 1993; McClelland et al. 1995; Rahman et al. 2008b). RAPD markers remained popular till the fall of 20th century in searching polymorphisms at inter and intra specific level for calculating genetic diversity and developing genetic linkage maps. Use of RAPD markers in genetic mapping is limited because of its low reliability and low stringency of annealing, and inability to check differences in sequence homology among similarly sized fragments (Jones et al. 1997). Moreover, the utility of RAPD assays in phylogenetic studies among the distantly related species or genera is a major concern (Pan et al. 1997) because markers are non-locus specific (Kesseli et al. 1994; Karp et al. 1996).

To overcome the reproducibility concern, these markers have been converted into SCARs for making them reliable and co-dominant in expression. However, RAPDs are not likely to regain popularity because of their low informativeness, and also the availability of ESTs and genome sequence information for help in designing primers for flanking regions of microsatellites to survey for polymorphisms.

4 Amplified Fragment Length Polymorphism (AFLP) in Cotton

4.1 Basis of the Method

A DNA fingerprinting technique called AFLP combines the reliability of RFLP with the convenience of PCR (Vos et al. 1995), generating reproducible genotyping profiles for complex genomes of any origin. The procedure avoids the need for prior sequence information similar to RAPD and contrary to SSRs (Table 2). It involves three steps: (i) restriction of the genomic DNA and ligation of oligonucleotide adapters; (ii) pre and selective amplification of sets of restriction fragments; and (iii) gel analysis of the amplified fragments. Typically 50–100 restriction fragments are amplified and detected on denaturing polyacrylamide gels. The polymorphic fragments are detected as present or absent, making it a dominant marker system like RAPD. AFLP can be automated using its modified and fluorescent version (fAFLP). Another advantageous feature is simultaneous analysis of many genetic loci per experiment. AFLP produces more polymorphic loci per primer than RFLPs, SSRs or RAPDs (Maughan et al. 1996).

The AFLP technique has been extensively used in DNA marker enrichment of specific regions to bridge gaps between genetic and physical maps (Vos et al. 1995); genome mapping studies, for example, in *Oryza* (Zhu et al. 1998), *Zea* (Xu et al. 1999) and *Solanum* (Bradshaw et al. 1998); population genetics and phylogeny analysis (Abdalla et al. 2001); and cultivar/accession identification (Iqbal et al. 2001); and MAS.

4.2 Application to Phylogenetic and Diversity Studies

AFLP markers were used to establish the extent of genetic diversity and relatedness in cotton. The AFLP method produces a 10-fold increase in the number of DNA fragments per plant, compared with RAPD methods. The AFLP data assigned the species–genotypes into groups that corresponded with their origin and/or pedigree relationships (Pillay and Myers 1999).

Evolutionary and genetic relationships of various germplasm resources including 43 cultivars of *G. raimondii*, *G. incanum*, *G. herbaceum* and *G. arboreum* were estimated using AFLPs (Iqbal et al. 2001). Genetic similarities suggested by AFLP analysis were in agreement with taxonomic relationships at the species level suggested by other groups using different marker systems (Abdalla et al. 2001; Murtaza 2006). *G. hirsutum* germplasm accessions showed exceedingly limited genetic diversity indicating that these accessions were derived from the same gene pool (Bligh et al. 1999; Pillay and Myers 1999). However, the genetic diversity is comparatively high among the *Gossypium* species (Abdalla et al. 2001).

4.3 Application to Linkage and QTL Mapping

A map based on F₂ population developed from a cross *G. hirsutum* acc. TM-1 x *G. barbadense* acc. 3–79 was constructed using RAPD and AFLP markers, comprising 11 linkage groups that covered 521.7 cM (Altaf et al. 1997). In another study, 490 AFLP markers associated with agronomic traits were identified using the F₂ population developed from an interspecific cross (Reddy et al. 1997). A backcross interspecific population was surveyed using 465 AFLP loci along with 229 SSRs, 192 RFLPs, and 2 morphological markers which resulted into 37 linkage groups covered 4400 cM distance (Lecape et al. 2003). Recently, more than 50 AFLP markers have been surveyed on 92 recombinant inbred lines (RILs) of *G. hirsutum* grown in replicated plots in China and the USA, and identified AFLPs associated with fiber and agronomic traits. However, one to four markers were associated with 22–93% of the phenotypic variability of each of the seven traits which suggest that the selected markers could be used in MAS (Jixiang et al. 2007).

Markers assigned to chromosomes are more useful than unlinked markers in MAS and map based cloning (Baogong 2004). Out of 42 linkage groups developed using an interspecific F₂ population, 19 were assigned to 12 chromosomes using aneuploid interspecific hybrids and a set of 29 RFLP and SSR framework markers. Seven QTLs were also detected for six fiber-related traits; five of these were distributed among A-subgenome chromosomes, the genome donor of fiber traits (Mei et al. 2004).

To identify abundant polymorphisms for mapping, a trispecific F₂ mapping population was screened with AFLP and RAPD markers. A linkage map containing 51 linkage groups spanning about 6,663 cM was developed exhibiting a higher level of recombination and polymorphism in the D genome than the A genome (Khan et al. 1998).

The possibility of identifying AFLPs as diagnostic markers for *G. hirsutum* and its closest relative *G. tomentosum* (endemic to the Hawaiian archipelago) was explored in a study where 11 and 16 species-specific markers were identified for *G. tomentosum* and *G. hirsutum*, respectively. These species-specific AFLP markers would be useful for detecting gene flow between *G. hirsutum* and *G. tomentosum* that had occurred in the past and might occur in the future (Hawkins et al. 2005).

4.4 Cleaved AFLP (cAFLP) – Enriching Polymorphism

The occurrence of AFLPs is relatively low in cotton (*G. hirsutum* or *G. barbadense*), limiting its utility in developing genome-wide linkage maps. In cAFLP, frequent use of restriction enzymes to cleave the AFLP amplicons increases the number of polymorphic markers (Zhang et al. 2005) by 67% and 132% in *G. hirsutum* and *G. barbadense* genotypes, respectively. Generally, both AFLP

and cAFLP procedures are similar; however, cAFLP-based genetic similarity (GS) is generally lower than AFLP-based GS. In some cases, cAFLP offers relatively better resolution than AFLP. Moreover, AFLP and cAFLP fragments amplified with the same primer combination can be pooled as one sample before electrophoresis if high resolution capillary-based DNA electrophoresis is used. Using several restriction enzyme combinations before pre-selective amplification in combination with various 4 bp-cutters or 6 bp-cutters after selective amplification, the pooled AFLP and cAFLP can provide large number of polymorphic markers for genome-wide mapping and fingerprinting.

4.5 AFLP-RGA – Identifying Candidate Resistance Genes

AFLP is also being used in combination with other marker systems to detect abundant polymorphisms in the genus *Gossypium*. Disease resistance (*R*) genes constituting the largest *R* gene family with domains of nucleotide binding sites (NBS) and leucine-rich repeats (LRR) have been utilised for designing degenerate primers. These degenerate primers are used in combination with one selective AFLP primer in PCR; the procedure is called AFLP-RGA (Hayes and Saghai-Marooof 2000). AFLP-RGA was employed in cotton to search for polymorphisms in putative RGAs (Zhang et al. 2007). The level of polymorphism detected with this technique was similar to that of AFLP. Out of 446 AFLP-RGA fragments amplified, 76 (17.0%) and 37 (8.3%) were polymorphic within four *G. hirsutum* L. genotypes and four *G. barbadense* L. cotton genotypes, respectively. The number of polymorphic AFLP-RGA fragments (256) between *G. hirsutum* and *G. barbadense* was much higher (57.4%). However, these markers could not differentiate several genotypes. Similarly, six DNA markers including two AFLP markers linked with resistance to fusarium wilt disease, (controlled by single gene *FWRI*) were identified in Sicala V-1. Degenerate primers were designed based on conserved motifs of *R* genes which are used in combination with AFLP. Approximately 300 polymorphic AFLP-RGA markers were identified, many of which were placed on an existing linkage map (Niu et al. 2006).

AFLP-RGA offers a means to search for RGAs genome-wide. Due to the distribution of RGAs or RGA clusters in the genome, genome-wide AFLP-RGA analysis provides a useful resource for mapping of *R* genes for disease resistance in cotton.

4.6 Scanning for Methylation Polymorphism

DNA methylation plays a vital role in epigenetic regulation of genes (Kato et al. 2003; Meng et al. 2003; Berger 2004; Chan et al. 2004) and predominantly occurs at CG and CNG symmetric sequences and also at nonsymmetric sites

(Gruenbaum et al. 1981; McClelland 1983; Oakeley and Jost 1996). Levels and patterns of methylation polymorphism (MP) at homologous loci in cotton was studied using a methylation-sensitive AFLP screen of MP diversity at CCGG sites, within *G. hirsutum* (Keyte et al. 2006). A set of 20 *G. hirsutum* accessions with known information on genetic polymorphism levels and relationships showed that MP exists at high levels within *G. hirsutum*. Out of the 150 HpaII/MspI sites surveyed, 48 were methylated at the inner cytosine (32%) and 32 of these were polymorphic (67%). Both these values are higher than estimates of genetic diversity obtained using RFLPs.

4.7 Summary – Strengths and Weaknesses of the Method

AFLPs have been extensively surveyed for developing genome maps, phylogenetic analysis, population genetics and cultivar/accession identification. Their expression is dominant and avoid the need for prior sequence information which is similar to RAPD but contrary to SSRs. Moreover, the AFLP procedure produces more polymorphic loci per experiment than RAPD and SSR. Like SSR, the AFLP system can be automated. The procedure is somewhat prone to experimental errors like RAPD, and its cost and time required are hindrances to its wider acceptance by the scientific community.

5 Simple-Sequence Repeat (SSRs) Based Methods

5.1 Basis of the Method

SSRs also known as microsatellites or short tandem repeats, are short 2–8 nucleotide motifs, repeated in tandem for a few to hundreds of times at many independent loci in eukaryotic genomes. SSRs are highly polymorphic (Lagercrantz et al. 1993; Powell et al. 1996). The variation in SSR arrays is thought to be due to either slippage of DNA polymerase during replication or unequal crossing over, resulting in differences in copy number of the core nucleotide sequences (Rahman et al. 2002a). Deletion or insertion of a single base or even long DNA fragments in the flanking regions is also a source of variation in SSRs (Buteler et al. 1999). SSR polymorphism also occurs because of structural rearrangement at the priming site and presence or absence of introgressed DNA in the amplified region (Paran and Michelmore 1993). Usually the sequences flanking the repeats are conserved between the members of a gene pool.

SSRs can be surveyed like RFLPs using a short synthetic oligonucleotide probe. However, to expedite the process, SSRs can be converted into a PCR-based marker system which requires sequence information for the flanking regions and the design of locus-specific primers. SSRs are easier to use than RFLPs, requiring a smaller amount of DNA and being more compatible with

automation (Mitchell et al. 1997). SSR markers can easily be exchanged among researchers because of their locus specificity. SSRs are abundant and widely distributed throughout the genomes of many higher plants and animals. These are now replacing RFLPs in genetic mapping of crop plants (Brondani et al. 1998; Korzun et al. 1999; Temnykh et al. 2000). Being based on longer primers, SSR assays are more robust than RAPDs and AFLPs (Table 2). The co-dominant nature of SSRs is also an advantage for genetic mapping (Han et al. 2006).

Mapping SSRs is useful in cotton because of the occurrence of one micro-satellite on average per 170 kb of genomic DNA (Zhao et al. 1994). SSRs are also present in expressed sequence tags (ESTs) (Saha et al. 2003) and dense consensus maps including EST-SSRs are possible functional genomic tools (He et al. 2007).

Capturing SSRs is accelerated by the use of enriched SSR libraries made from genomic DNA, which facilitates high throughput screening and automated DNA sequencing (Connell et al. 1998). Out of 10,000 SSR containing genomic fragments isolated from *G. hirsutum*, 588 were sequenced to identify SSRs and primers were designed for 307 of these (Reddy et al. 2001). Among a subset of SSR markers screened on *G. hirsutum* and *G. barbadense* varieties, approximately 49% were polymorphic, while lower polymorphism of ~25–30% between *G. hirsutum* var FH-1000 and *G. barbadense* acc PGMB-36 was reported (Mumtaz 2007). SSR isolation from genomic DNA randomly sheared with nitrogen gas ensures unbiased representation of the genome (Kumapatla et al. 2004).

Most SSRs identified to date in the public sector have been summarized in the “cotton microsatellite database” (CMD), which provides information for about 5,484 cotton SSRs, publications, sequences, primers, mapping and homology data for nine cotton microsatellite projects (Blenda et al. 2006).

Multiplexing of SSRs is possible as several SSRs can be amplified in a single reaction rather than pooling PCR products from numerous individual reactions before electrophoretic separation on a gel (Mitchell et al. 1997; Liu et al. 2000). Through multiplexing, 55 SSRs were successfully used in genome mapping of a fertility restorer gene in cotton (Liu et al. 2003) which would improve the efficiency of genetic mapping and marker-assisted programs (Liu et al. 2000).

5.2 Application to Phylogenetic and Diversity Studies

Knowledge about the phylogenetic relationships among cotton species is important to understand their evolutionary relationships (Khan et al. 2000; Abdalla et al. 2001; Paterson et al. 2002), estimate the extent of genetic diversity (Khan et al. 2000) and plan breeding strategies (Rahman et al. 2002b). Some marker systems applied in cotton are handicapped because of low heterozygosity detected using isozymes and RFLPs (Brubaker and Wendel 2000), and importantly lack of reproducibility of RAPD markers (Jones et al. 1997).

The genus *Gossypium* has been the focus of many researchers over the last two decades, offering opportunities for understanding genome evolution, plant development and productivity (Rong et al. 2004; Guo et al. 2005). A high level of variation was detected using SSRs in 10 *G. hirsutum* cultivars collected from diverse U.S. production regions when compared with RFLPs. Specifically, an average of 1.64 different DNA fingerprints per primer pair was obtained, while RFLPs yielded an average of 1.28 different DNA fingerprints per probe (Rong et al. 2004). In a similar study, an average of 1.48 DNA fingerprints per primer pair was reported in *G. arboreum* genotypes collected from diverse regions of Pakistan (Yasmin 2005).

Evolution of homeologous SSRs is independent in polyploid species, and the homeologous sequences maintained their identity after different genomes were joined into one nucleus (Cronn et al. 1999; Liu and Wendel 2001). Homeologous sequences of many SSRs in polyploids are phylogenetic sisters, detected in tetraploid cotton species (Guo et al. 2005).

For tetraploid cotton, some reports find the level of allelic diversity in SSRs of the D-subgenome to be higher than A-subgenome (Small et al. 1999; Adams and Wendel 2004). However, other studies found comparable allelic diversity in both genomes (Lecape et al. 2007). Similarly, the nucleotide diversity of the *Adh* gene family in the D-subgenome was 2–3 times higher than for the A-subgenome (Small et al. 1999). Such commonalities were also reported for wheat (Thuillet et al. 2004).

SSRs surveyed on races and wild species of *Gossypium* exhibited significantly higher levels of polymorphism than within cultivated *G. hirsutum* (Gutierrez et al. 2002; He et al. 2007; Lecape et al. 2007) which are congruent with earlier results obtained using allozymes as well as DNA markers (Brubaker and Wendel 1994; Iqbal et al. 2001; Rahman et al. 2002a; Rungis et al. 2005; Shaheen 2005). Similar results were reported using SSRs on *G. arboreum* germplasm (Rahman et al. 2006b). It was suggested that allelic richness in the cultivated pool of *G. hirsutum* (Abdalla et al. 2001) and *G. arboreum* (Rahman et al. 2006b) represents a small fraction of their respective species variability (Abdalla et al. 2001). However, most cultivars from geographically isolated regions could be uniquely fingerprinted using a small number of SSRs (Gutierrez et al. 2002; Rahman et al. 2002b; Shaheen 2005). Genetic diversity and F_2 bulk performance were compared upon surveying 90 SSRs on selected cotton genotypes bred in different countries (five U.S. and four Australian cultivars, and two day-neutral converted lines of *G. hirsutum*). A weak association was found between low estimates of genetic diversity and F_2 -bulk population performance. However, if the genotypes were diverse, then genetic diversity was the better predictor for some agronomic and fiber traits (Gutierrez et al. 2002).

SSR diversity in cotton and its related species (Liu et al. 2000; Rungis et al. 2005; Rahman et al. 2006b), was found to be low within modern annualized cultivars (Lecape et al. 2007), a major handicap in their utilization (Iqbal et al. 2001; Rungis et al. 2005) especially in plant breeding programs (Gupta and Varshney 2000). Limited genetic variation in cotton is, perhaps, due to its

narrow genetic base (Liu et al. 2000; Iqbal et al. 2001) which resulted from attempting intraspecific crosses (Lu and Myers, 2002; Rahman et al. 2002b) or backcrossing of the adapted germplasm.

Ninety five Pakistani cotton cultivars/genotypes including six Chinese genotypes, two US and one Indian genotype of *G. hirsutum* were surveyed with 50 SSRs including 25 EST-SSRs, out of which 24 (48%) were polymorphic, revealing a low level of polymorphism. The SSRs were found more polymorphic than the EST-SSRs. In this study, most of the varieties released in the pre-CLCuD era made one major cluster, and the cultivars/varieties released in post-CLCuD era made a second major cluster. However, the cultivars of Chinese origin formed a third cluster (Shaheen 2005; Rahman et al. 2006b).

The EST-SSRs isolated from *G. hirsutum* successfully amplified the corresponding loci from other cotton species (Liu et al. 2000; Qureshi et al. 2004; Mumtaz 2007). The ability to amplify these markers by PCR, conservation of sequence across different species of cotton and a high polymorphism rate make the EST-SSRs ideal for comparative mapping in cotton (Qureshi et al. 2004).

Most SSR primer pairs in cotton amplified multiple products, as the result of multiple priming sites along the genome (Fisher et al. 1998; Rallo et al. 2000; Saal et al. 2001) which are particularly common in allopolyploids (Guo et al. 2005). Often the patterns obtained are complex due to competition and the co-migrating bands (Rallo et al. 2000).

5.3 Application to Linkage Mapping

Generally, cotton genetic maps constructed from interspecific populations (Reinisch et al. 1994; Jiang et al. 1998; Kohel et al. 2001; Saranga et al. 2001; Zhang et al. 2002; Paterson et al. 2003; Han et al. 2004; Mei et al. 2004; Nguyen et al. 2004; Rong et al. 2004; He et al. 2005; Lacape et al. 2005; Park et al. 2005; Song et al. 2005; Frelichowski et al. 2006; Han et al. 2006; Hua et al. 2007) exhibited more genome coverage than maps developed from intraspecific populations (Shappley et al. 1998; Ulloa and Meredith 2000; Zhang et al. 2003; Shen et al. 2005, 2007; Ulloa et al. 2005).

At present, about $\frac{1}{4}$ of the 5484 available cotton SSRs (<http://www.mainlab.clemson.edu/cmd/>) have been mapped on *G. hirsutum*/*G. barbadense* genotypes to identify DNA markers associated with QTLs for drought tolerance, yield components and fiber properties (Liu et al. 2000; Nguyen et al. 2004; Han et al. 2004, 2006; Park et al. 2005). One reason that “The International Cotton Genome Initiative (ICGI)” was launched, was to facilitate the development of a saturated and fully integrated genetic and physical map of cotton that will pave the way for developing a consensus linkage map using a common set of framework markers, such as SSRs (Yu et al. 2005).

5.4 Application to Gene Tagging and QTL Mapping

A first report describing the association of SSRs with plant resistance gene provoked thoughts in identifying diagnostic SSRs in cotton associating with various traits of interest (Table 1; Rahman 2002; Nguyen et al. 2004). Out of the 250 SSRs surveyed on *G. hirsutum*, two markers were associated with genes conferring resistance to CLCuD (Rahman 2002; Rahman et al. 2005a) which were further utilized in monitoring the transfer of these genes in succeeding generations (Rahman et al. 2006a), paved the way for developing two cotton lines NIBGE-2 (Rahman et al. 2007b) and NIBGE-115 (Rahman et al. 2007a) which are extremely resistant to an old strain of CLCuD and extremely tolerant to a newly evolved resistant breaking 'Burewala' virus strain. Similarly, a study to detect markers linked to the verticillium wilt resistance genes in *G. hirsutum* using BSA, assembled 35 polymorphic SSRs into 11 linkage groups spanning 531 cM with an average distance of 15.17 cM (Bolek et al. 2005). An inter-specific $F_{2:3}$ population derived from a cross between *G. hirsutum* cv XLZ1 (susceptible cultivar) and *G. barbadense* cv H-7124 was genotyped with SSRs which mapped into 41 linkage groups spanning around 3745.9 cM. A total of nine QTLs conferring resistance against the disease explained 10.63–28.83% of the phenotypic variance, out of these six were located on the D sub-genome (Wang et al. 2008).

Hairiness and red leaf color traits involved in defense against insect pests, were mapped in *G. hirsutum* by surveying 54 SSRs. Of these, two SSRs (JESPR-153 and JESPR-300) were linked to hairiness and three SSRs (CM-43, CM-162 and JESPR-204) were associated with leaf color. The preliminary genetic linkage map for hairiness showed that the SSR JESPR-153 had a map distance of 15.7 cM from the hairiness locus. Two SSRs, CM-162 and JESPR-204, were mapped at 5 cM and 19.4 cM, respectively, from the red leaf color locus (Ali 2004).

For developing a map to identify SSRs linked with fiber QTLs, an F_2 population developed from a cross between a lintless mutant and TM-1 was surveyed with 22 polymorphic SSRs, located on eight linkage groups covering 218.3 cM. Two SSRs of the *Lil* (Ligon lintless) linkage group were located on chromosome 22 (Karaca et al. 2002). In another study, 489 loci (out of 510 SSRs and 114 RAPDs) assembled into 43 linkage groups covering 3,314.5 cM. The linkage groups were assigned to chromosomes using monosomic and telodisomic genetic stocks (Zhang et al. 2002). Similarly, an interspecific *G. hirsutum* x *G. barbadense* backcross population comprising 75 BC_1 plants was surveyed with 1014 markers. The map consisted of 888 loci including 465 AFLPs, 229 SSRs, 192 RFLPs, and 2 morphological markers, ordered in 37 linkage groups spanning 4400 cM. The loci were not evenly distributed over the linkage groups, and 18 of the 26 groups had a single dense region (Lacape et al. 2003). Two SSRs and six RAPDs were found associated with two QTLs for fiber strength using interspecific F_2 and F_3 populations derived from a cross between

G. anomalum introgression line 7235 (known for fiber quality) and TM-1 (Zhang et al. 2003). Similarly, a genetic map containing 1,160 SSRs covered 5,519 cM, with an average distance between two loci of 4.8 cM which can be utilized for developing a consensus map and for the effective monitoring of QTLs (Nguyen et al. 2004).

In QTL mapping experiments, a number of QTLs associated with fiber length were identified including one QTL (Akash 2003), one QTL (Mei et al. 2004), 15 QTLs (Lacape et al. 2005), one QTL (Park et al. 2005), five QTLs (Frelichowski et al. 2006), eight QTLs (Shen et al. 2007) and five QTLs (Hua et al. 2007). Similarly, two QTLs (Akash 2003), one QTL (Mei et al. 2004), five QTLs (Lacape et al. 2005), two QTLs (Park et al. 2005), five QTLs (Frelichowski et al. 2006), five QTLs (Shen et al. 2007) and eight QTLs (Hua et al. 2007) associated with fiber fineness were reported. Also the QTLs associated with fiber strength including three QTLs explaining 47.1% PV (Asif 2007), three QTLs (Hua et al. 2007), four QTLs (Kohel et al. 2001), one major QTL (Zhang et al. 2003), two QTLs (Mei et al. 2004), 12 QTLs (Lacape et al. 2005), two QTLs (Park et al. 2005), six QTLs (Frelichowski et al. 2006) and seven QTLs (Shen et al. 2007) were identified.

5.5 Inter-Simple Sequence Repeat (ISSR) Markers

A modification of SSR-based markers, called ISSR involves the use of micro-satellite-complementary oligonucleotides as primers (repeats of 1–3 bases), circumventing the requirement for flanking sequence information (Wolfe et al. 1998) (Table 2). The procedure was first described and applied to fungi (Meyer et al. 1993) and then to animals (Perring et al. 1993), and plants (Gupta et al. 1994) for finding polymorphisms. In addition to freedom from the necessity of obtaining flanking genomic sequence information, ISSR analysis is technically simpler than many other marker systems. The method provides highly reproducible results and generates abundant polymorphisms in many systems. The amplicons can be size fractionated on agarose gel stained with ethidium bromide (Nagaoka and Ogihara 1997; Joshi et al. 2000) or on polyacrylamide gel employing silver staining, or using isotopic detection (Blair et al. 1999).

ISSRs have been surveyed in cotton using sequencing gels, and fluorescent labelling of amplification primers, and found to be an easy and informative genetic marker system in cotton for revealing both inter- and intraspecific polymorphisms (Liu and Wendel 2001). ISSR analysis is flexible in designing experiments, optimizing the number of amplicons by modifying the primer design (e.g. the number of anchored bases) and primer combinations. ISSR analysis has been made user friendly using fluorescein-labelling of amplification primers and automated detection on sequencing gels (Huang and Sun 2000).

5.6 Summary – Strengths and Weaknesses of the Methods

SSRs are highly polymorphic (Lagercrantz et al. 1993; Powell et al. 1996) and reproducible (Jones et al. 1997), and their expression is co-dominant (Han et al. 2006). SSRs are PCR-based and compatible with automation, making them markers of choice. Unlike RAPDs and AFLPs, prior knowledge of the flanking regions containing SSRs is prerequisite for designing primers, however, such information is now more easily and cheaply obtained than in the past. Mapping of SSR containing ESTs isolated from various cDNA libraries would be instrumental in identification and or isolation of cotton genes (Paterson et al. 2000; Han et al. 2006; Udall et al. 2006). The mapping studies can be further accelerated through developing community resources such as integrated Web database (Gingle et al. 2006) and cotton microsatellite database (CMD) (Blenda et al. 2006). In the future, numerous resources including a pilot sequence of cotton being generated by an international consortium under the leadership of Prof Andrew Paterson, Univ Georgia will be helpful in identifying new SSRs, which will pave the way for initiating MAS in cotton improvement programs (Chen et al. 2007).

6 Single Nucleotide Polymorphisms (SNPs) – the Future of Cotton DNA Markers

Genetic improvement of cotton fiber production and its properties will be enhanced by the availability of rapidly developing genetic resources and tools, including high-density genetic maps (Rong et al. 2004; Lacape et al. 2005) (Table 2). SNPs are single base change or small insertions and deletion (Indels) in homologous DNA fragments. SNPs are the most abundant source of polymorphisms. For example, human genome contains ~9–10 million SNPs, of which 3.1 million have been identified (The International Hapmap consortium 2007), and present in coding as well as non-coding regions of genomes (Aerts et al. 2002). As a marker, SNPs are co-dominant in dosage, distributed normally and sometimes associated with morphological changes (Lindblad-Toh et al. 2000), thus making them superior to other marker systems. SNPs are useful for characterizing allelic variation, QTL mapping, and implementing MAS. In *Arabidopsis*, more than 37,000 SNPs were identified in earlier studies (Jander et al. 2002; Schmid et al. 2003). Recently, new re-sequencing approaches (array-based methods) have been developed to identify SNPs. For example, in *Arabidopsis*, more than one million non redundant SNPs were identified which are the excellent genetic resource for disequilibrium mapping studies (Clark et al. 2007). Similarly, higher plants like barley (Kanazin et al. 2002), maize (Tenailon et al. 2001), soybean (Zhu et al. 2003) and sugar beet (Schneider et al. 2001) have been surveyed for discovering SNPs.

Among the SNP detection techniques, one approach is to sequence the whole genome of any two individuals followed by comparison of loci in different individuals. It is difficult to identify SNPs in cotton because of huge genome size coupled with the polyploid nature of cultivated cotton, requiring the distinction of allelic SNPs from paralogs. Data mining has been done throughout the diploid genomes of cotton by direct sequencing of PCR products amplified by primers designed on STS or already mapped RFLPs localized on the chromosome. Minimal rate of variation per nucleotide (1.06%) between the two subgenomes (At and Dt) of allotetraploid cotton has been reported which is substantially higher than the level of variations between species (0.35%) or genotypes within species (0.37 and 0.14%) (Rong et al. 2004).

SSR primers can be used to discover SNPs in the flanking regions of SSRs in cotton (Shah et al. 2004). DNA sequencing of 21 genes involved in fiber development revealed heterogeneity of nucleotide diversity in two tetraploid cultivated cotton species, using a PCR based direct sequencing technique. A total of 94 SNPs including 36 single-base changes (38.3%) and 58 indels (61.7%) were identified in 16 fiber gene fragments (Lu et al. 2005). The average frequency of SNPs is approximately one SNP per 500 bp DNA in the selected fiber genes, which is lower than that in coding sequences of many other plant species. In the *FIFI* gene, regulating the fiber development in *G. barbadense*, three base substitutions were reported while comparing with the corresponding gene in *G. hirsutum* (Ahmad et al. 2007). In another study, in six *R2R3-MYB* transcription factors influencing trichome length and density, one SNP per 77 bases were reported (An et al. 2008).

In *G. hirsutum*, the frequency of SNPs is lower than in *G. barbadense*. SNPs exhibited an unequal distribution between and within genes, with more indels than base substitution. ESTs derived from *G. hirsutum* and its A and D genome progenitors were examined for alloallelic SNPs, resulting in 2342 orthologous gene pairs distinguished by 10,000 SNPs and indels (Udall et al. 2006). In another study, only a limited number of SNPs were found in chloroplast genes. For example, nucleotide differences in the *rps8* gene of *G. arboreum* with the corresponding gene sequences were in the range of 0% for *G. hirsutum* to 17.41% for *Epiphagus virginiana* (Shaheen et al. 2006).

ESTs have also been exploited for mining and developing SNPs in combination with AFLP technology to mitigate the complexity of the polyploid genome. For *G. hirsutum* using selective AFLP primer combinations or one AFLP primer in combination with one gene targeted primer, DNA or cDNA libraries were constructed. SNPs were detected by comparing the genotypes after plasmid DNA sequencing. This AFLP-based SNP strategy allows high-throughput, low cost detection of SNP markers in cotton (Zhang et al. 2007).

Sequencing is informative and useful in providing a starting point to understand how genes vary in different populations, but obtaining information on multiple individuals in a population is expensive. Several alternatives to sequencing have been proposed to discover allelic polymorphisms (Kwok 2001; Tsuchihashi and Dracopoli 2002). Analysis of DNA for single stranded

conformational polymorphism (SSCP) provides an efficient measure to screen for point mutations before the costly and time consuming task of sequencing is begun. Under specific electrophoretic conditions SSCP helps to identify mutations by means of conformational changes elicited through differences in migration patterns (Atha et al. 1998). A very familiar SNP and mutation discovery method is mass spectrometry (MS), and Sequenom (San Diego, CA, USA) developed a very effective large-scale genotyping method (Buetow et al. 2001). The mutation detection technology used in Targeting Induced Local Lesions in Genomes (TILLING) has been adapted to the discovery of polymorphism in natural populations (Comai et al. 2004). Another effective technology is pyrosequencing, which measures the release of a pyrophosphate with addition of a base during each round of DNA synthesis (Langae and Ronaghi 2005).

SNP genotyping techniques are being improved rapidly. Various methods have been developed based on allele-specific hybridization, primer extension, oligonucleotide ligation, endonuclease cleavage or allelic specific PCR (Gupta et al. 2001; Gut 2001; Shi 2001; Syvanen 2001). Genotyping methods using fluorescence energy transfer have also been developed, requiring specialized detection equipment with a flow cytometric platform and modified fluorescence-labeled primers (Livak 1999; Cai et al. 2000; Chen et al. 2000; Taylor et al. 2001; Ye et al. 2001). These methods are more suitable than the conventional detection methods to multiplex large number of markers for genetic mapping and diversity studies (Kim et al. 2005). Microarray based SNP genotyping can be a very effective tool but it is just in preliminary stages. These methods can improve the pace of genotyping in cotton.

References

- Abdalla, A.M., Reddy, O. U. K., El-Zik, K. M. and Pepper, A. E. (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor. Appl. Genet.* 102 (2/3), 222–229.
- Adams, K.L. and Wendel, J.F. (2004) Exploring the mysteries of polyploidy in cotton. *Biol. J. Linn. Soc.* 82, 573–581.
- Aerts, J., Wetzels, Y., Cohen, N. and Aerssens, J. (2002) Data mining of public SNP database for the selection of intragenic SNPs. *Hum. Mutat.* 20, 162–173.
- Ahmad, M., Khan, M.R. and Saeed, M. (1987) Studies on factors contributing resistance in five new cultivars of cotton against insect pests. *Pak. Entomol.* 9 (1–2), 23–28.
- Ahmad, S., Zhang, T., Noor-ul-Islam, Shaheen, T. and Rahman, M. (2007) Identifying genetic variation in *Gossypium* based on single nucleotide polymorphism. *Pak. J. Bot.* 39(4), 1245–1250.
- Akash, M. W. (2003) Quantitative trait loci mapping for agronomic and fiber quality traits in upland cotton (*Gossypium hirsutum* L.) using molecular markers. PhD dissertation. Louisiana State Univ. USA.
- Ali, I. (2004) Development of genetic linkage maps for some insect non-preference traits in cotton. PhD Thesis, Univ. Agric. Faisalabad Pakistan.
- Ali, I., Ashraf, M., Rahman, M., Ullah, I., Niaz, M., Hanif, M., Hassan, M. and Zafar, Y. (2004) Screening of cotton varieties regarding hairiness trait for DNA fingerprinting. *J. Nat. Sci* 2(1), 91–99.

- Altam, M.K., Stewart, J.M., Wajahatullah, M.K. and Zhang, J.F. (1997) Molecular and morphological genetics of a trispecies F₂ population of cotton. In: P. Dugger, and D.A. Richter (Eds.), *Proceedings of the Beltwide Cotton Production Research Conferences*. National Cotton Council of America, Memphis, TN. Animal Genome VI, Jan 18–22. 1998. San Diego California, USA, pp. 448–452.
- An, C., Saha, S., Jenkins, J.N., Ma, D-P., Scheffler, B.E., Kohel, R.J., Yu, J.Z., Stelly, D.M. (2008) Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping. *Theor. Appl. Genet.* 116, 1015–1026.
- Asif, M., Rahman, M. and Zafar, Y. (2005) DNA fingerprinting studies of some wheat (*Triticum aestivum* L.) genotypes using random amplified polymorphic DNA (RAPD) analysis. *Pak. J. Bot.* 37(2), 271–277.
- Asif, M. (2007) Genomic Analysis for Quality Traits in Cotton (*Gossypium hirsutum* L.) by DNA Fingerprinting Technology. PhD Thesis, B.Z.Univ. Multan Pakistan.
- Asif, M., Rahman, M. and Zafar, Y. (2006) Genotyping Analysis of Maize (*Zea mays* L.). Hybrids Using DNA Fingerprinting Technology. *Pak. J. Bot.* 38 (5), 1425–1430.
- Aslam, M., Jiang, C., Wright, R. and Paterson, A.H. (1999) Identification of molecular markers linked to leaf curl virus disease resistance in cotton. *Pak. J. Bio. Sci.* 2(1), 124–126.
- Atha, D.H., Wenz, H.M., Morehead, H., Tian, J. and O'Connell, C.D. (1998) Detection of p53 point mutation by single strand conformational polymorphism (SSCP): analysis by capillary electrophoresis. *Electrophoresis* 19, 172–179.
- Baogong, J. (2004) Optimization of agrobacterium mediated cotton transformation using shoot apices explants and quantitative trait loci analysis of yield and yield component traits in upland cotton (*Gossypium hirsutum*). PhD thesis. Louisiana State Univ.
- Bassam, B.J., Caetano-Anollés, G. and Gresshoff, P.M. (1995) Method for profiling nucleic acids of unknown sequence using arbitrary oligonucleotide primers. US Patent 5, 413–909.
- Becelaere, G.V., Lubbers, E.L., Paterson, A.H. and Chee, P.W. (2005) Pedigree- vs. DNA Marker-Based Genetic Similarity Estimates in Cotton. *Crop Sci.* 45:2281–2287.
- Beckmann, J.S. and Soller, M. (1983) Restriction length polymorphisms in general improvement: methodologies, mapping and cost. *Theor. Appl. Genet.* 67, 35–43.
- Berger, F. (2004) Imprinting—a green variation. *Science* 303, 483–485.
- Bhatnagar, P. and Sharma P.D. (1991) Comparative incidence of sucking insect pests on different isogenic lines of cotton variety. H777. *Insect Sci.* 4, 170–171.
- Blair, M.W., Panaud, O. and McCouch, S.R. (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L). *Theor. Appl. Genet.* 98, 780–792.
- Blenda, A., Scheffler, J., Scheffler, B., Palmer, M., Lacape J.M., Yu, J.Z., Jesudurai, C., Jung, S., Muthukumar, S., Yellambalase, P., Ficklin, S., Staton, M., Eshelman, R., Ulloa, M., Saha, S., Burr, B., Liu, S., Zhang, T., Fang, D., Pepper, A., Kumpatla, S., Jacobs, J., Tomkins, J., Cantrell, R. and Main, D. (2006) The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *B.M.C. Genomics* 7, 61.
- Bligh, H. F. J., Blackhall, N. W., Edwards, K. J. and McClung, A. M. (1999) Using amplified fragment length polymorphism and simple sequence length polymorphism to identify cultivars of brown and white milled rice. *Crop Sci.* 39, 1715–1721.
- Bolek, Y., El-Zik, K.M., Pepper, A.E., Bell, A.A., Magill, C.W., Thaxton, P.M. and Reddy, O.U. (2005) Mapping of verticillium wilt resistance genes in cotton. *Plant Sci.* 168, 1581–1590.
- Botstein, D., White, R. L., Skolnik, M. and Davis, R. W. (1980) Construction of a genetic linkage map in using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Bowman, D.T. (1999) Public cotton breeders – do we need them? *J. Cotton Sci.* 3, 139–152.

- Bowman, D.T., May, O.L. and Calhoun, D.S. (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Sci.* 36, 577–581.
- Bradshaw, J.E., Hackett, C.A., Meyer, R.C., Milbourne, D., McNichol, J.W., Philips, M.S. and Waugh, R. (1998) Identification of AFLP and SSR marker associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. *Theor. Appl. Genet.* 97, 202–210.
- Brondani, R.P.V., Brondani, C., Tarchini, R. and Grattapaglia, D. (1998) Development, Characterisation and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor. Appl. Genet.* 97, 816–827.
- Brubaker, C.L. and Wendel, J.F. (1994) Re-evaluating the origin of domesticated cotton using nuclear restriction fragment length polymorphism. *Am. J. Bot.* 81, 1309–1326.
- Brubaker, C.L., Paterson, A.H. and Wendel, J.F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42, 184–203.
- Brubaker, C.L. and Wendel, J.F. (2000) RFLP diversity in cotton. In: J.N. Jenkins and S. Saha (Eds.), *Genetic Improvement of Cotton: Emerging Technologies*. Sci. Publishers, Inc. USA, pp. 81–102.
- Buetow, K.H., Edmonson, M. and MacDonald, R. (2001) High throughput development and characterization of a genome wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci.* 98, 581–584.
- Buteler, M.I., Jarret, R.L. and LaBonte, D. R. (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. *Theor. Appl. Genet.* 99, 123–132.
- Butter, N. S. and Vir, B. K. (1991) Response of whitefly *Bemisia tabaci* (Genn.) to different cotton genotypes under glass house conditions. *Ind. J. Entomol.* 53 (1), 115–119.
- Caetano-Anollés, G., Bassam, B.J. and Gresshoff, P.M. (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/technol.* 9, 553–557.
- Caetano-Anollés, G. (1996) Scanning of nucleic acids by *in vitro* amplification: new developments and applications. *Nat. Biotech.* 14, 1668–1674.
- Cai, H., Whites, P.S., Torney, D., Dishpande, A., Wang, Z., Marrone, B. and Nolan, J.P. (2000) Flow cytometry-based minisequencing: A new platform for high output single nucleotide polymorphism scoring. *Genomics* 66, 135–142.
- Cantrell, R.G., Pederson, J. and Liu, S. (1999) Mapping of introgressed cotton populations with DNA markers. *Plant & Animal Genome VII Conference*, January 17–21, San Diego, CA, USA.
- Chan, S.W.L., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C. and Jacobsen, S.E. (2004) RNA silencing genes control de novo DNA methylation. *Science* 303, 336.
- Chee, P.W., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, B., Smith, C.W. and Paterson, A.H. (2005a) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. *Theor. Appl. Genet.* 111, 757–763.
- Chee, P.W., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, B., Smith, C.W. and Paterson, A.H. (2005b) Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. *Theor. Appl. Genet.* 111, 772–781.
- Chee, P., Jiang, C., Draye, X., Decanini, L.I., Smith, C.W., DelMonte, T.A. and Paterson, A. H. (2000) Advanced-backcross mapping of cotton QTLs. *Plant & Animal Genome VIII Conference*, San Diego, USA.
- Chen, J., Ennone, M.A., Li, M.S., Tylor, J.D., Rivers, P., Nelson, A.J., Slentz-Casler, K., Roses, A. and Weiner, M.P. (2000) A microsphere based assay for multiplex single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* 10, 549–557.

- Chen, Z.J., Scheffler, B.E., Dennis, E., Triplett, B., Zhang, T., Chen, X., Stelly, D.M., Rabinowicz, P.D., Town, P., Arioli, T., Brubaker, C., Cantrell, R., Lacape, J., Ulloa, M., Chee, P., Gingle, A.R., Haigler, C.H., Percy, R., Saha, S., Wilkins, T., Wright, R.J., Van Deynze, A., Zhu, Y., Yu, X., Guo, W., Abdurakhmonov, I., Katageri, I., Rahman, M., Zafar, Y., Yu, J.Z., Kohel, R.J., Wendel, J. and Paterson, A.H. (2007) Towards sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 145, 1251–1263.
- Cherry, J.P., Katterman, F.R.H. and Endrizzi, J.E. (1970) Comparative studies of seed proteins of species of *Gossypium* by gel electrophoresis. *Evolution* 24, 431–447.
- Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H., Frazer, K.A., Huson, D.H., Schölkopf, B., Nordborg, M., Rätsch, G., Ecker, J.R. and Weigel D. (2007) Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*. *Science* 317, 338–342.
- Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Jhonson, J.E., Burtner, C., Odden, A.R. and Henikoff, S. (2004) Efficient discovery of DNA polymorphism in natural populations by Ecotilling. *Plant J.* 37, 778–786.
- Connell, J.P., Pammi, S., Iqbal, M.J., Huizinga, T. and Reddy, A.S. (1998) A high throughput procedure for capturing microsatellites from complex plant genomes. *Plant Mol. Biol. Rep.* 16, 341–349.
- Cronn, R.C., Small, R.L., Haselkorn, T. and Wendel, J.F. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* 89, 707–725.
- Cronn, R., Small, R.L. and Wendel J.F. (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl. Acad. Sci.* 96, 14406–14411.
- Cronn, R.C., Zhao, X.P., Paterson, A.H., Wendel, J.F. (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705.
- Draye, X., Chee, P., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, R., Smith, C.W. and Paterson, A.H. (2005). Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach. II Fiber Fineness. *Theor. Appl. Genet.* 111, 764–771.
- Dudley, J.W., Saghai-Maroo, A. and Rufener, G.K. (1992) Molecular marker information and selection of parents in corn breeding programs. *Crop Sci.* 32, 301–304.
- Feng, C.D., Stewart, J.McD. and Zhang, J.F. (2005) STS markers linked to the Rf1 fertility restorer gene of cotton. *Theor. Appl. Genet.* 110, 237–243.
- Fisher, P. J., Richardson, T. E. and Gardner, R. C. (1998) Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. *Theor. Appl. Genet.* 96, 969–979.
- Frelichowski, J.E.Jr., Palmer, M.B., Main, D., Tomkins, J.P., Cantrell, R.G., Stelly, D.M., Yu, J., Kohel, R.J. and Ulloa, M. (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. *Mol. Genet. Gen.* 275(5), 479–491.
- Fryxell, P.A. (1971) Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (*Malvaceae*). *Evolution* 25, 554–562.
- Giese, H., Holm-Jensen, A.G. and Jensen, J. (1993) Localization of the *laevigatum* powdery mildew resistance gene to barley chromosome 2 by the use of RFLP markers. *Theor. Appl. Genet.* 85, 897–900.
- Gingle, A.R., Yang, H., Chee, P.W., May, O.L., Rong, J., Bowman, D.T., Lubbers, E.L., Day, J.L. and Paterson, A.H. (2006) An Integrated Web Resource for Cotton. *Crop Sci.* 46, 1998–2007.
- Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) Sequence specificity of methylation in higher plant DNA. *Nature* 292, 860–862.
- Guo, W., Zhang, T., Shen, X., Yu, J.Z. and Kohel, R.J. (2003) Development of SCAR marker linked to a major QTL for high fiber strength and its usage in molecular-marker assisted selection in upland cotton. *Crop Sci.* 43(6), 2252–2257.

- Guo, W.Z., Fang, D., Yu, W.D. and Zhang, T.Z. (2005) Sequence divergence of microsatellites and phylogeny analysis in tetraploid cotton species and their putative diploid ancestors. *J. Integ. Plant Biol.* 47(12), 1418–1430.
- Gupta, P.K., Roy, J.K. and Parsad, A. (2001) Single nucleotide polymorphism; A paradigm for molecular marker technology and DNA polymorphism detection with emphasis with their use in plants. *Curr. Sci.* 80, 524–535.
- Gupta, M., Chyi, Y.S., Romero-Severson J. and Owen J.L. (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* 89, 998–1006.
- Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185.
- Gut, I.J. (2001) Automation in genotyping of single nucleotide polymorphism. *Hum. Mutat.* 46, 416–419.
- Gutierrez, O.A., Basu, S., Saha, S., Jenkins, J.N., Shoemaker, D.B., Cheatham, C.L. and McCarty, J.C. (2002) Genetic distance among selected cotton genotypes and its relationship with F_2 performance. *Crop Sci.* 42, 1841–1847.
- Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) A novel repeat element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci.* 79, 6465–6469.
- Hayes, A.J. and Saghai-Marooof, M.A. (2000) Targeted resistance gene mapping in soybean using modified AFLPs. *Theor. Appl. Genet.* 100, 1279–1283.
- Han, Z.G., Guo, W., Song, X.L. and Zhang, T. (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Mol. Genet. Gen.* 272, 308–327.
- Han, Z.G., Wang, C., Song, X.L., Guo, W., Gou, J.Y., Li, C., Chen, X. and Zhang, T. (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor. Appl. Genet.* 112, 430–439.
- Hawkins, J. S., Pleasants, J. and Wendel, J.F. (2005) Identification of AFLP markers that discriminate between cultivated cotton and the Hawaiian island endemic, *Gossypium tomentosum* Nuttall ex Seeman. *Genet. Resour. Crop Evol.* 52, 1069–1078.
- He, D.H., Zhong, X.L., Zhang, X. L., Nie, Y.C., Guo, X.P., Feng, C.D. and Stewart, J.McD. (2005) Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. *Euphytica* 144, 141–149.
- He, D.H., Lin, Z.X., Zhang, X.L., Nie, Y.C., Guo, X.P., Zhang, Y.X. and Li, W. (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. *Euphytica* 153 (1–2), 181–197.
- Helentjaris, T., Slocum, M., Wright, S., Schaefer, A. and Nienhuis, J. (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphism. *Theor. Appl. Genet.* 72, 761–769.
- Hua, H.D., Xu, L.Z., Long, Z.X., Chun, N.Y., Ping, G.X., Xin, Z.Y. and Wu, L. (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. *Euphytica* 153 (1–2), 181–197.
- Huang, J. and Sun, S.M. (2000) Genetic diversity and relationships of sweet potato and its wild relatives in *Ipomoea* series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet.* 100, 1050–1060.
- Iqbal, M.J., Aziz, N., Saeed, N.A., Zafar, Y. and Malik, K.A. (1997) Genetic diversity of some elite cotton varieties by RAPD analysis. *Theor. Appl. Genet.* 94, 139–144.
- Iqbal, M.J., Reddy, O.U.K., El-Zik, K.M. and Pepper, A.E. (2001) A genetic bottleneck in the evolution under domestication of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* 103, 547–554.

- Jander, G., Norris, S., Rounsley, S., Bush, D., Levin, I. and Last, R. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* 129: 440–450.
- Jiang, C.X., Wright, R. J., El-Zik, K. M. and Paterson, A. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci.* 95, 4419–4424.
- Jiang, C.X., Chee, P.W., Draye, X., Morrell, P.L., Smith, C.W. and Paterson, A.H. (2000) Multi-locus interactions restrict gene flow in advanced-generation interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54, 798–814.
- Jixiang, W., Johnie, N.J., McCarty, J.C., Zhong, M. and Michael, S. (2007) AFLP marker associations with agronomic and fiber traits in cotton. *Euphytica* 153, 153–163.
- Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini P., Buiatti, M., Maestri, E., Malcevski, A., Marmioli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., and Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* 3, 381–390.
- Johnson, B.L. and Thein, M.M. (1970) Assessment of evolutionary affinities in *Gossypium* by protein electrophoresis. *Amer. J. Bot.* 57, 1081–1092.
- Joshi, S.P., Gupta, V.S., Aggarwal, R.K., Ranjekar, P.K. and Brar, D.S. (2000) Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theor. Appl. Genet.* 100, 1311–1320.
- Kanazin, V., Talbert, H., See, D., DeCamp, P., Nevo, E. and Blake, D. (2002) Discovery and assay of single nucleotide polymorphism in barley (*Hordeum vulgare*). *Plant Mol. Biol.* 48, 529–537.
- Kato, M., Miura, A., Bender, J., Jacobsen, S.E. and Kakutani, T. (2003) Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Biol.* 13, 421–426.
- Karaca, M., Saha, S., Jenkins, J.N., Zipf, A., Kohel, R. and Stelly, D.M. (2002) Simple sequence repeat (SSR) markers linked to the Ligon lintless (Li (1)) mutant in cotton. *J. Heredity* 93(3), 221–224.
- Karp, A., Seberg, O. and Buiatti, M. (1996) Molecular techniques in the assessment of botanical diversity. *Ann. Bot.* 78, 143–149.
- Kesseli, R., Paran, I. and Michelmore, R.W. (1994) Analysis of a detail linkage map of *Lactuca sativa* (Lettuce) constructed from RFLP and RAPD markers. *Genetics* 136, 1435–1446.
- Keyte, L.A., Percifield, R., Liu, B. and Wendel, J.F. (2006) Intraspecific DNA Methylation Polymorphism in Cotton (*Gossypium hirsutum* L.). *J. Heredity* 97(5), 444–450.
- Khan, M.A., Zhang, J. and Stewart, J.McD. (1998) Integrated molecular map based on a trispecific F₂ population of cotton. *Proceeding Beltwide cotton improvement conference* San Diego, CA. January 5–9.
- Khan, S. A., Hussain, D., Askari, E., Stewart, J. McD., Malik, K.A. and Zafar, Y. (2000) Molecular phylogeny of *Gossypium* species by DNA fingerprinting. *Theor. Appl. Genet.* 101(5/6), 931–938.
- Kim, M.Y., Van, K., Lestari, P., Moon, J.K. and Lee, S.H. (2005) SNP identification and SNAP marker development for a *GmNARK* gene controlling super nodulation in soybean. *Theor. Appl. Genet.* 110, 1003–1010.
- Klein-Lankhorst, R.M., Vermunt, A., Weide, R., Liharska, T. and Zabel, P. (1991) Isolation of molecular markers for tomato (*Lycopersicon esculentum*) using random amplified polymorphic DNA (RAPD). *Theor. Appl. Genet.* 83, 108–114.
- Kohel, R.J., Yu, J. and Park, Y.H. (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica* 121, 163–172.
- Knight, R.L. and Sadd, J. (1954) The genetics of jassid resistance in cotton. III. The Kapas Puro, Kawanda and Philippines Ferguson group. *J. Genet.* 52, 186–198.

- Korzun, V., Rtjder, M.S., Wendehake, K., Pasqualone, A., Lotti, C., Ganai, M.W. and Blanco, A. (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat, *Theor. Appl. Genet.* 98, 1202–1207.
- Kumpatla, S.P., Manley, M.K., Horne, E.C., Gupta, M. and Thompson, S.A. (2004) An improved enrichment procedure to develop multiple repeat classes of cotton microsatellite markers. *Plant Mol. Biol. Rep.* 22: 85a–85i
- Kwok, P.Y. (2001) Methods for genotyping single nucleotide polymorphism. *Annu. Rev. Gen. Hum. Genet.* 2, 235–258.
- Lacape, J. M., Nguyen, T.B., Courtois, B., Belot, J.L., Giband, M., Gourlot, J.P. Gawryziak, G., Roques, S. and Hau, B. (2005) QTL Analysis of Cotton Fiber Quality Using Multiple *Gossypium hirsutum* x *Gossypium barbadense* Backcross Generations. *Crop Sci.* 45,123–140.
- Lacape, J.M., Nguyen, T.B., Thibivilliers, S., Bojinov, B., Courtois, B., Cantrell, R.G., Burr, B. and Hau, B. (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* 46, 612–626.
- Lacape, J.M., Dessauw, D., Rajab, M., Noyer, J.L. and Hau, B. (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol. Breed.* 19, 45–58.
- Lagercrantz, U., Ellegren, H. and Andersson, L. (1993) The abundance of various polymorphic microsatellite motifs differ between plants and vertebrates. *Nucl. Acids Res.* 21, 1111–1115.
- Lan, T.H., Cook, C.G. and Paterson, A.H. (1999) Identification of a RAPD marker linked to a male fertility restoration gene in cotton (*Gossypium hirsutum* L.). *J. Agric. Gen.* 4:1–5.
- Langae, T. and Ronaghi, M. (2005) Genetic variation analyses by Pyrosequencing. *Mutat. Res.* 573, 96–102.
- Lazo, G. R., Park, Y. and Kohel, R. J. (1994) Identification of RAPD markers linked to fiber strength in *Gossypium hirsutum* and *G. barbadense* interspecific crosses. *Proceeding Biochemistry of Cotton*. September 28–30, 1994. Galveston, Texas.
- Lindblad-Toh, K., Wibchester, E., Daly, M.J., Wang, D.G., Hirschhof, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklae, P., Shah, N., Thomas, D., Fan, J.B., Grigeras, T., Warrington, J., Patil, N., Hudson, T.J. and Lander, E.S. (2000) Large-scale discovery and genotyping of single nucleotide polymorphism in mouse. *Nat. Genet.* 24, 381–386.
- Liu, S., Saha, S., Stelly, D., Burr, B. and Cantrell, R.G. (2000) Chromosomal assignment of microsatellite loci in cotton. *J. Heredity* 91(4), 326–32.
- Liu, B. and Wendel, J.F. (2001) Intersimple sequence repeat (ISSR) polymorphisms as a genetic marker system in cotton. *Mol. Ecol. Notes* 1, 205–208.
- Liu, L., Guo, W., Zhu, X. and Zhang, T. (2003) Inheritance and fine mapping of fertility restoration for cytoplasmic male sterility in *Gossypium hirsutum* L. *Theor. Appl. Genet.* 106, 461–469.
- Livak, K.J. (1999) Allelic discrimination using flourogenic probe and 5nuclease assay. *Genet. Anal.* 14, 143–149.
- Lu, Y., Curtiss, J., Zhang, J., Percy, R.G. and Cantrell, R.G. (2005) Discovery of single nucleotide polymorphisms in selected fiber genes in cultivated tetraploid cotton. *National Cotton Council Beltwide Cotton Conference*, pp. 946.
- Maughan, P.J., Saghai Maroof, M.A., Buss, G.R. and Huestis, G.M. (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance and near-isogenic line analysis. *Theor. Appl. Genet.* 93:392–401.
- McClelland, M. (1983) The frequency and distribution of methylatable DNA sequences in leguminous plant protein coding genes. *J. Mol. Evol.* 19, 346–354.
- McClelland, M., Mathieu-Daude, F. and Welsh, J. (1995) RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends in Genet.* 11, 242–246.

- Mei, M., Syed, N. H., Gao, W., Thaxton, P.M., Smith, C.W., Stelly, D.M. and Chen, Z.J. (2004) Genetic mapping and QTL analysis of fiber related traits in cotton (*Gossypium*). *Theor. Appl. Genet.* 108, 280–291.
- Meng, L., Bregitzer, P., Zhang, S. and Lemaux, P.G. (2003) Methylation of the exon/intron region in the ubil promoter complex correlates with transgene silencing in barley. *Plant Mol. Biol.* 53,327–340.
- Meredith, W.R.J. (1992) Improving fiber strength through genetics and breeding. In: C.R. Benedict and G.M. Jividen (Eds.), *Proceeding Cotton Fiber Cellulose: Structure, Function, and Utilization Conference*, Savannah, GA. 28-31 Oct. 1992. National Cotton Council, Memphis, TN, pp. 289–302.
- Meyer, W., Mitchell, T.G., Freedman, E.Z. and Vilgalys, R. (1993) Hybridization probes for conventional DNA fingerprinting can be used as single primers in the PCR to distinguish strains of *Cryptococcus neoformans*. *J. Clin. Microbiol.* 31, 2274–2280.
- Mitchell, S.E., Kresovich, S., Jester, C.A., Hernandez, C.J. and Szewc-McFadden, A.K. (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Sci.* 37,617–624
- Mukhtar, M.S., Rahman, M. and Zafar, Y. (2002) Assessment of genetic diversity among wheat (*Triticum aestivum* L.) cultivars from a range of localities across Pakistan using random amplified polymorphic DNA (RAPD) analysis. *Euphytica* 128, 417–425.
- Multani, D.S. and Lyon, B.R. (1995) Genetic Fingerprinting of Australian cotton cultivars with RAPD markers. *Genome* 38, 1005–1008.
- Mumtaz, H. (2007) Identification of structural and functional genomic markers for fiber quality traits in cotton using interspecific population (*G. hirsutum* x *G. barbadense*). MPhil Thesis, QA Univ Islamabad Pakistan.
- Murtaza, N. (2006) Cotton genetic diversity study by AFLP markers. *Electronic J. Biotech.* 9 (4), 456–460.
- Nagaoka, T. and Ogihara, Y. (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet.* 94, 597–602.
- Nguyen, T.B., Giband, M., Brottier, P., Risterucci, A.M. and Lacape, J.M. (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109, 167–175.
- Niu, C., Lu, Y. and Zhang, J. (2006) Plant disease resistance gene analogues in cotton: mapping and expression. *Beltwide Cotton Conference*, Jan 3-6. 2006 San Antonio Texas.
- Oakeley, E.J. and Jost, J.P. (1996) Non-symmetrical cytosine methylation in tobacco pollen DNA. *Plant Mol. Biol.* 31, 927–930.
- Pan, Y.B., Burner, D.M., Ehrlich, K.C., Grisham, M.P. and Wei, Q. (1997) Analysis of primer-derived, non-specific amplification products in RAPD-PCR. *Biotech.* 22, 1071–1077.
- Paran, I. and Micheltore, R.W. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* 85, 985–993.
- Parks, C.R., Ezell, W.I., William, D.E. and Dreyer, D.L. (1975) The application of flavonoid distribution to taxonomic problems in the genus *Gossypium*. *Bulletin of the Torrey Botanical Club* 102, 350–361.
- Park, Y.H., Alabady, M.S., Ulloa, M., Sickler, B., Wilkins, T.A., Yu, J., Stelly, D.M., Kohel, R.J., El-Shihy, O.M. and Cantrell, R.G. (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol. Genet. Gen.* 274, 428–441
- Parnell, F.R., King, H.E. and Ruston, D.F. (1949) Jassid resistance and hairness of cotton plant. *Bull. Entomol. Res.* 39, 539–575.
- Paterson, A.H., Damon, S. Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S. and Tanksley S., (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127, 181–197.

- Paterson, A.H., Brubaker, C.L. and Wendel, J.F. (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genome DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* 11, 122–127.
- Paterson, A.H. and Zhao, X. (1995) Molecular tools for cotton improvement. In Herber D (Ed). *Proceeding Beltwide Cotton Production Research Conference* San Antonio, TX, 4–7 Jan., National Cotton Council of Amer., Memphis, TN.
- Paterson, A.H., Bowers, J., Burrow, M., Draye, X., Elisk, C., Jiang, C., Katsar, C., Lan, T., Lin, Y. and Ming, R. (2000) Comparative genomics of plant chromosomes. *Plant Cell* 12, 1523–1539.
- Paterson, A., Estill, J., Rong, J., Williams, D. and Marler, B. (2002) Toward a genetically anchored physical map of the cotton genomes. *Cotton Sci.* 14 (Suppl.), 31.
- Paterson, A.H., Saranga, Y., Menz, M., Jiang, C.X. and Wright, R.J. (2003) QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theor. Appl. Genet.* 106, 384–396.
- Perring, T.M., Copper, A.D., Rodriguez, R.J., Farrarand, C.A. and Bellows, T.S. (1993) Identification of a white fly species by genomics and behavioral studies. *Science* 259, 74–77.
- Phillips, L.L. (1966) The cytology and phylogenetic of the diploid species of *Gossypium*. *Ann. J. Bot.* 53, 328–335.
- Pillay, M. and Myers, G.O. (1999) Genetic Diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Sci.* 39, 1881–1886.
- Powell, W., Machray, G.C. and Provan, J. (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1, 215–222.
- Qureshi, S.N., Saha, S., Kantety, R.V. and Jenkins, J.N. (2004) EST-SSR: a new class of genetic markers in cotton. *J. Cotton Sci.* 8, 112–123.
- Rafalski, J.A. and Tingey, S.V. (1993) Genetic diagnostics in plant breeding: RAPDs, micro-satellites and machines. *Trends in Genet.* 9, 275–279.
- Rahman, M. (2002) DNA markers studies in cotton. PhD Thesis, Department of Plant Breed. Genet. Univ. Agric. Faisalabad Pakistan.
- Rahman, M., Ahmed, N., Asif, M. and Zafar, Y. (2006^a) Identification of DNA markers linked with cotton leaf curl disease (CLCD). *International Cotton Genome Initiative (ICGI) Workshop*, 18–20 Sept Brasilia Brazil. Pp. 77–78.
- Rahman, M., Asif, M., Ali, I., Hayder, Z., Malik, K.A. and Zafar, Y. (2003). DNA marker studies in cotton. *Final meeting of CFC/ICAC 07 project* held at ICBA, Dubai 28–30 September 2003.
- Rahman, M., Asif, M., Ullah, I., Malik, K.A. and Zafar, Y. (2005)^a Overview of cotton genomic studies in Pakistan. *Plant & Animal Genome Conference XIII*. San Diego, CA. USA.
- Rahman, M., Aslam, N., Asif, M., Malik, T.A., Malik, K.A. and Zafar, Y. (2002^a) Identification of DNA markers for cotton leaf curl disease (CLCD) in cotton (*Gossypium hirsutum* L.). *Cotton Sci.* 14, 17.
- Rahman, M., Hussain, D. and Zafar, Y. (2002^b) Estimation of genetic divergence among elite cotton cultivars–genotypes by DNA fingerprinting technology. *Crop Sci.* 42, 2137–2144.
- Rahman, M., Malik, T.A., Hussain, D. and Zafar, Y. (2005)^b Genetics of resistance to cotton leaf curl virus disease in *Gossypium hirsutum*. *Plant Pathol.* 54, 764–772.
- Rahman, M., Malik, T.A., Iqbal, M.J., Zafar, Y. and Malik, K.A. (2004) Gene tagging for salinity resistance in wheat by RAPD. *Pak. J. Bot.* 36 (3), 595–602.
- Rahman, M., Ullah, I., Ashraf, M. and Zafar, Y. (2008^a) A study of genotypic variation for drought tolerance in cotton. *Agron. Sustain. Dev.* 28, 439–447.
- Rahman, M., Yasmin, T., Shaheen, H.L., Shaheen, T., Tabassam, N. and Zafar, Y. (2006^b) Harvesting genetic diversity in the cotton genome. *National symposium on biotechnology for economic prosperity*, National Commission on Biotechnology, Green Retreat Hotel, Nathiagali, Pakistan, July 24–26, 2006. pp 38.

- Rahman, M., Yasmin, T., Tabassum, N., Ullah, I., Asif, M. and Zafar, Y. (2008^b) Studying the extent of genetic diversity among *Gossypium arboreum* L. genotypes/cultivars using DNA fingerprinting. *Genetic Res. Crop Evo.* 55, 331–339.
- Rahman, M. and Zafar, Y. (2001) Genotyping of a new strain of lentil (*Lens Culinaris* Medik.) by DNA fingerprinting technology. *Pak. J. Bot.* 33 (4), 423–428.
- Rahman, M. and Zafar, Y. (2007^a). Registration of NIBGE-115. *Journal of Plant Registrations*, 1, 51–52.
- Rahman, M. and Zafar, Y. (2007^b) Registration of NIBGE-2. *Int. J. Plant Registration* 2, 113–114.
- Rallo, P., Dorado, G. and Martin, A. (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). *Theor. Appl. Genet.* 101, 984–989.
- Rana, M.K. and Bhat, K.V. (2004) A Comparison of AFLP and RAPD Markers for Genetic Diversity and Cultivar Identification in Cotton. *J. Plant Biochem. Biotech.* 13, 19–24.
- Reddy, A., Haisler, R.M., Yu, J. and Kohel, R.J. (1997) AFLP mapping in cotton. *Plant Animal Genome Conference V*, San Diego CA, USA.
- Reddy, O.U.K., Pepper, A.E., Abdurakhmonov, I., Saha, S., Jenkins, J.N., Brooks, T.B., Bolek, Y. and El-Zik, K.M. (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *J. Cotton Sci.* 5, 103–113.
- Reinisch, A.J., Dong, J.M., Brubaker, C.L., Wendel, J.F. and Paterson, A.H. (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138, 829–847.
- Rong, J.K., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C., Pierce, G.J., Rainey, K.M., Rastogi, V., Schulze, K., Tronlinde, N.L., Wendel, J.F., Wilkins, T.A., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. and Paterson, A.H. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166, 389–417.
- Rungis, D., Llewellyn, D., Dennis, E.S. and Lyon, B.R. (2005) Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Austr. J. Agri. Res.* 56, 301–307.
- Shah, M.R., Kumpatla, S.P., Gupta, M. and Thompson, S.A. (2004) SSR-Based SNP discovery: an economical approach for the development of SNP markers in cotton. *Plant & Animal Genomes XII Conference*, San Diego, USA.
- Saha, S., Karaca, M., Jenkins, J.N., Zipf, A.E., Reddy, O.U.K. and Kantety, R.V. (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. *Euphytica* 130(3), 355–364.
- Saal, B., Plieske, J., Hu, J., Quiros, C.F. and Struss, D. (2001) Microsatellite markers for genome analysis in *Brassica*. II. Assignment of rapeseed microsatellites to the A and C genomes and genetic mapping in *Brassica oleracea* L. *Theor. Appl. Genet.* 102 (5), 695–699.
- Saranga, Y., Menz, M., Jiang, C.X., Wright, R.J., Yakir, D. and Paterson, A.H. (2001) Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. *Genome Res.* 11, 1988–1995.
- Schmid, K.J., Sørensens T.R., Stracke, R., "rjék, O.T., Altmann, T., Mitchell-Olds, T. and Weisshaar B. (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res* 13:1250–1257.
- Schneider, K., Weisshaar, B., Borchardt, D.C. and Salamini, F. (2001) SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. *Mol. Breed.* 8, 63–74.
- Schon, C.C., Lee, M., Melchinger, A.E., Guthrie, W.D. and Woodman, W.L. (1993) Mapping and characterization of quantitative trait loci affecting resistance against second generation European corn borer in maize with the aid of RFLPs. *Heredity* 70, 648–659.

- Shaheen, H.L. (2005) Global view of genetic diversity among cotton cultivars/genotypes by microsatellite analysis. MPhil Thesis. Department of Botany Univ. Agric. Faisalabad Pakistan.
- Shaheen, T., Rahman, M. and Zafar, Y. (2006) Chloroplast RPS8 gene of cotton reveals the conserved nature through out taxa. *Pak J Bot.* 38 (5): 1467–1476.
- Shappley, Z.W., Jenkins, J.N., Watson, C.E., Kahler, A.L. and Meredith, W.R. (1996) Establishment of molecular markers and linkage groups in two F₂ populations of upland cotton. *Theor. Appl. Genet.* 92, 915–919.
- Shappley, Z.W., Jenkins, J.N., Zhu, J. and McCarty, J.C. (1998) Quantitative trait loci associated with agronomic and fiber traits of Upland Cotton. *J. Cotton Sci.* 4, 153–163.
- Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J.Z., Kohel, R.J. and Zhang, T. (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol. Breed.* 15 (2), 169–181.
- Shen, X., Guo, W., Lu, Q., Zhu, X., Yuan, Y. and Zhang, T. (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. *Euphytica* 155, 371–380.
- Shi, M.M. (2001) Enabling large scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* 47, 164–172.
- Small, R.L., Ryburn, J.A. and Wendel, J.F. (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* 16, 491–501.
- Song, X. L., Wang, K., Guo, W.Z., Zhang, J. and Zhang, T.Z. (2005) A comparison of genetic maps constructed from haploid and BC₁ mapping populations from the same crossing between *Gossypium hirsutum* L. 3 *G. barbadense* L. *Genome* 48, 378–390.
- Struss, D. and Plieske, J. (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. *Theor. Appl. Genet.* 97, 308–315.
- Syvanen, A.C. (2001) Assessing genetic variation genotyping single nucleotide polymorphism. *Nat. Rev. Genet.* 2, 930–942.
- Tabassam, N., Rahman, M. and Zafar Y. (2006) DNA-based genotyping of sorghum hybrids. *Pak. J. Bot.* 38, 1599–1604.
- Tatineni, V., Cantrell, R.G. and Davis, D.D. (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPD. *Crop Sci.* 36, 186–192.
- Taylor, G.D., Briley, D., Nguyen, Q., Long, K., Iannone, M.A., Li, M.S., Ye, F., Ashari, A., Lai, A., Wagner, M., Chen, J. and Weiner, M.P. (2001) Flow cytometric platforms for high-throughput single nucleotide polymorphism analysis. *Biotech.* 30, 661–669.
- Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovick, L., Cho, Y.G., Ishii, T. and McCouch, S.R. (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100, 697–712.
- Tenaillon, M., Sawkins, M., Long, A., Gaut, R., Doebley, J. and Gaut, B. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays*). *Proc. Natl. Acad. Sci.* 98, 9161–9166.
- Tsuchihashi, Z. and Dracopoli, N.C. (2002) Progress in high throughput SNP genotyping methods. *Pharmacogenomics J.* 2, 103–110.
- The International HapMap Consortium. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 451, 451–461.
- Thuillet, A.C., Bataillon, T., Sourdille, P. and David, J.L. (2004) Factors affecting polymorphism at microsatellite loci in bread wheat [*Triticum aestivum* (L.) Thell]: effects of mutation processes and physical distance from the centromere. *Theor. Appl. Genet.* 108(2), 368–377.
- Udall, J.A., Quijada, P.A., Lambert, B. and Osborn, T.C. (2006) Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor. Appl. Genet.* 113, 597–609.

- Ulloa, M. and Meredith, W.R. (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *J. Cotton Sci.* 4, 161–170.
- Ulloa, M., Meredith, W.R., Shappley, Z.W. and Kahler, A.L. (2002) RFLP genetic linkage maps from four $F_{2:3}$ populations and a joinmap of *Gossypium hirsutum* L. *Theor. Appl. Genet.* 104, 200–208.
- Ulloa, M., Saha, S., Jenkins, J.N., Meredith, W.R., Mccarty, J.J.C., and Stelly, D.M. (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) Joinmap. *J. Heredity* 96(2), 132–144.
- Valicek, P. (1978) Wild and cultivated cottons. *Cottoned fibres tropics* (English Eds.) 33, 363–387.
- Van Esbroeck, G.A., Bowman, D.T., May, O.L and Calhoun, D.S. (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. *Crop Sci.* 39, 323–328.
- Vos, P., Hogers, R., Blecker, M., Reijans, M., Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* 23, 4407–4414.
- Wang, G.L., Dong, J.M. and Paterson, A.H. (1995) The distribution of *Gossypium hirsutum* chromatin in *Gossypium barbadense* germplasm: molecular analysis of introgressive plant breeding. *Theor. Appl. Genet.* 91, 1153–1161.
- Wang, H-M., Lin, Z-X., Zhang, X-L., Chen, W., Guo, X-P., Nie, Y-C. and Li Y-H. (2008) Mapping and quantitative trait loci analysis of verticillium wilt resistance genes in cotton. *Journal of Integrative Plant Biol.* 50(2), 174–182.
- Welsh, J. and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18, 7213–7218.
- Wendel, J.F. and Albert, V.A. (1992) Phylogenetics of the cotton genus (*Gossypium*)-character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot.* 17, 115–143.
- Wendel, J.F., Brubaker, C.L. (1993) RFLP diversity in *Gossypium hirsutum* L. and new insights into the domestication of cotton. *Amer J Bot* 80: 71.
- Wendel, J.F., Brubaker, C.L. and Percival, A.E. (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* 79, 1291–1310.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18, 6531–6535.
- Wolfe, A.D., Xiang, Q.Y. and Kephart, S.R. (1998) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Mol. Ecol.* 7, 1107–1125.
- Wright, R.J., Thaxton, P.M., El-Zik, K.M. and Paterson, A.H. (1999) Molecular mapping of genes affecting pubescence of cotton. *J. Heredity* 90 (1), 215–219.
- Wright, R., Thaxton, P., Paterson, A.H. and El-Zik, K. (1998) Polyploid formation in *Gossypium* has created novel avenues for response to selection for disease resistance. *Genetics* 149, 1987–1996.
- Xu, M.L., Melchinger, A.E., Xia, X.C. and Lübberstedt, T. (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR and AFLP markers. *Mol. Gen. Genet.* 261, 574–581.
- Yasmin, T. (2005). DNA based characterisation of different cultivated cotton strains. MPhil Thesis, Department of Botany Univ. Agric. Faisalabad Pakistan.
- Ye, F., Li, M.S., Taylor, J.D., Nguyen, Q., Colton, H.M., Casey, W.M., Wagner, M., Weiner, M.P. and Chen, J. (2001) Fluorescent microsphere-based readout technology for multiple human single nucleotide polymorphism analysis and bacterial identification. *Hum. Mutat.* 17, 305–316.

- Yu, J., Park, Y., Lazo, G.R., Wolff, N.C. and Kohel, R.J. (1996) Molecular mapping of the cotton genome and its applications to cotton improvement. *Proceeding Beltwide Cotton Conference*, pp. 636.
- Yu, J., Kohel, R.J., Xu, Z., Dong, J., Zhang, H., Stelly, D.M., Zhu, Y. and Covalada, L. (2005) Physical mapping of fiber development genes in cotton. *Plant & Animal Genome XIII Conference*, Town & Country Hotel, San Diego, CA, pp. 225.
- Zhang, J. and Stewart, J.McD. (2004) Identification of molecular markers linked to the fertility restorer genes for CMS-D8 in cotton. *Crop Sci.* 44, 1209–1217.
- Zhang, J., Guo, W. and Zhang, T. (2002) Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L. x *Gossypium barbadense* L.) with a haploid population. *Theor. Appl. Genet.* 105 (8), 1166–1174.
- Zhang, J., Pang, M.X., Niu, C., Wang, W., Percy, R.G., Cantrell, R.G. and Stewart, J. McD. (2007) AFLP-based SNP discovery in cotton. *Plant & Animal Genome XV Conference*, Town & Country Hotel, San Diego, CA, pp 169.
- Zhang, J., Stewart, J.McD. and Wang, T. (2005) Linkage analysis between gametophytic restorer *Rf₂* gene and genetic markers in cotton. *Crop Sci.* 45, 147–156.
- Zhang, T.Z., Yuan, Y., Yu, J., Guo, W. and Kohel, R.J. (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. *Theor. Appl. Genet.* 106, 262–268.
- Zhang, Q. and Yu, S. (1999) Molecular markers-based gene tagging and its impact on rice improvement. In: J.S. Nanda (Ed.), *Rice Breeding and Genetics-Research Priorities and Challenges*, Science Publishers Inc., Enfield, New Hampshire, pp. 241–270.
- Zhang, J.F. and Stewart, J.McD. (2001). Inheritance and genetic relationships of the D₈ and D₂₋₂ restorer genes for cotton cytoplasmic male sterility. *Crop Sci.* 41, 289–294.
- Zhao, X.P., Lin, Y.R. and Paterson, A.H. (1994) Characterization and genetic mapping of DNA microsatellites from cotton. *Plant & Animal genome II conference*, Town & Country Conference Center, San Diego, CA.
- Zhu, J., Gale, M.D., Quarrie, S., Jackson, M.T. and Bryan, G.J. (1998) AFLP markers for the study of rice biodiversity. *Theor. Appl. Genet.* 96, 602–611.
- Zhu, Y.L., Song, Q.J., Hyten, D.L., Van, T.C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D. and Cregan, P.B. (2003) Single nucleotide polymorphism in soybean. *Genetics* 168, 1123–1134.

Physical Composition and Organization of the *Gossypium* Genomes

Lifeng Lin and Andrew H. Paterson

Abstract The 8 different diploid *Gossypium* genomes vary about three-fold in genome size. DNA renaturation kinetic analyses more than 30 years ago suggested that much of this variation was attributable to the repetitive DNA fractions, and subsequent cloning and sequencing studies have revealed specific DNA elements and families that contribute to this variation. The relationship between physical quantity of DNA and genetic distance (recombination fraction) in a region shows striking variation along individual *Gossypium* chromosomes, but an appreciable degree of correspondence across subgenomes and species due largely to conserved locations of centromeres. A substantial and growing collection of bacterial artificial chromosome (BAC) libraries for *Gossypium* species and genotypes provides a platform for studies of local organization of specific genomic regions, and for global physical characterization (which is in progress for several genomes). Of particular importance in planning for the sequencing of members of the *Gossypium* genus is the nearly two-fold difference in size between the A and D diploid genome types that have contributed to tetraploid cotton, and the finding that repetitive DNA has been transmitted between these two genomes (especially from A to D) in tetraploid cottons. Additional information being assembled, about the diversity among different members of the major repetitive element families, and the degree of inter-genomic exchange following polyploidization, will be important to devising cost-effective sequencing strategies.

1 Overview

From a common ancestor thought to have existed about 5-10 million years ago, the eight different diploid genome types in the *Gossypium* genus (distinguished from one another as described in Konan and Wendel chapters, this volume)

L. Lin (✉)
Plant Biology Department, University of Georgia
e-mail: phill@uga.edu

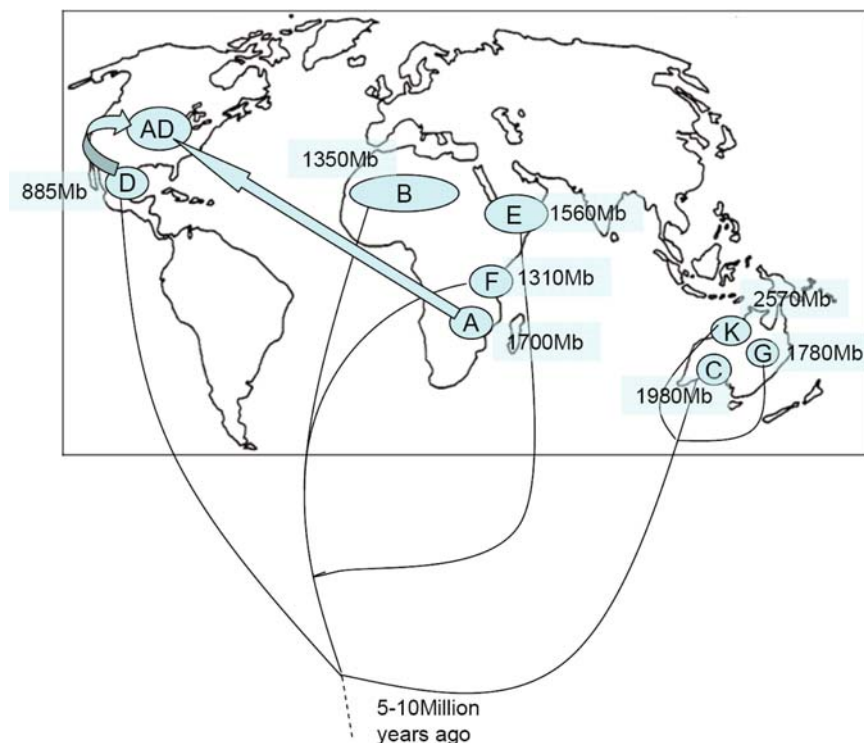


Fig. 1 The genome size and evolutionary relationship among different cotton species. Modified from <http://www.eeob.iastate.edu/faculty/WendelJ/images/map2.jpg> (See Color Insert)

have evolved striking differences in physical composition and nearly three-fold variation in genome size. The genome size and evolutionary relationships among the cotton species are shown in Fig.1. Among the diploid genomes, the K genome is the largest, with an estimated genome size even larger than that of the tetraploid (Wendel, Cronn et al. 2002). The D genome is the smallest, at about 900Mbp. All other cotton diploid genomes have genome sizes between 1300Mbp and 2000Mbp.

A variety of approaches have been employed in the dissection of the molecular basis of this variation. Cotton was an early subject of DNA reassociation kinetics studies, which yielded a general picture of cotton genome organization and comparative evolution of genome size that continues to be applicable today. However, newer methods have permitted us to dissect the 'kinetic components' of cotton DNA into individual DNA element families with different sequences, genomic distributions, and evolutionary strategies; to identify particular chromosomal regions in which there are striking deviations from the 'average' physical/genetic distance relationship; and to clone and characterize selected chromosomal segments. Repetitive DNA is the largest component of

eukaryotic genomes and is a key consideration in whole genome sequencing (Paterson 2006), therefore current and ongoing research in this area is important to designing cost-effective strategies by which to capture the unique sequence information that distinguishes the respective cotton genomes from one another and from those of other organisms.

2 Characterization of Cotton Genome Composition

2.1 Comparison of Cotton Genomes by DNA Reassociation Kinetics

In early efforts to analyze the DNA composition of the cotton genomes, quantitative measurements of the DNA content of the A, D and AD genomes were obtained using DNA reassociation kinetics, or ‘*C_{ot}* analysis’ (Kadir 1976; Walbot and Dure 1976; Geever, Katterman *et al.* 1989). In this procedure, genomic DNA is sheared into fragments and denatured, and then allowed to reassociate under controlled conditions with continuous monitoring of the portion of DNA that has renatured. DNA elements that are present in many (thousands of) copies in a genome renature rapidly, while elements present in few copies such as many genes renature slowly. In recent years this procedure has been used in conjunction with cloning to selectively clone and characterize DNA element families with differing abundance in a genome (Peterson, Schulze *et al.* 2002; Peterson, Wessler *et al.* 2002).

An early reassociation kinetic analysis of tetraploid cotton (*G. hirsutum*) provided our first glimpse into cotton genome composition (Walbot and Dure 1976). Highly repetitive DNA elements, with an average *C_{ot}* value of less than 0.1, comprised about 8% of the genome, and moderately repetitive elements with a *C_{ot}* value of 5.42 comprised about 27% of the genome. The remaining 60% of the tetraploid genome showed high *C_{ot}* values consistent with low copy number (excluding the small portion of DNA that is invariably damaged during such experiments).

C_{ot} analysis of the putative ancestors of the tetraploid cotton, the A1 (*G. herbaceum*) and D5 (*G. raimondii*) genomes (Geever, Katterman *et al.* 1989) revealed substantial differences in genome composition and organization. Specifically, the “zero time” (*C_{ot}* around 10^{-3} , extremely repetitive or self-annealing), moderately repetitive and single copy fragments comprise 7%, 54% and 39% of the A1 genome, respectively; and 7%, 30% and 63% of the D5 genome (Geever, Katterman *et al.* 1989), indicating that the D genome is substantially less repetitive than the A genome. The tetraploid genome (*G. hirsutum*) was re-evaluated, with 6%, 46% and 48% of the respective components, differing from Walbot and Dure’s estimate due to somewhat different circumscription of the three components. Notably, the tetraploid values (Geever, Katterman *et al.* 1989) are intermediate between those of the constituent A and D genome diploids, albeit somewhat closer to the A genome values. This is consistent

with the fact that roughly two-thirds of the tetraploid DNA is A-genome derived, in that the A genome contains roughly twice as much DNA as the D.

The sequence similarity between the A1 and D5 genome was estimated by reannealing of mixtures of DNA from the two species, comparing interspecific hybridization results with intraspecific hybridization to estimate the similarity between genomes. Reciprocal experiments showed 76.4% – 78.7% re-naturation, which we now know to be explicable by very high similarity of low-copy sequence, with appreciable divergence of many repetitive DNA families in the two genomes (see below).

2.2 Cloning and Characterization of Cotton Repetitive Element Families

DNA cloning permitted individual *Gossypium* repetitive elements to be isolated and studied. In a detailed characterization of repetitive elements in tetraploid cotton, a genomic library was screened by hybridization to labeled total genomic DNA, identifying 313 putatively repetitive clones that showed particularly strong hybridization signal. The clones were cross-hybridized to one another, and grouped into 103 families that differed in genome organization, methylation pattern, abundance, and DNA variation (Zhao, Wing *et al.* 1995). High abundance families were estimated by slot blot analysis to range from 15000 to 100000 copies, while moderate-abundance families ranged from 4000 to 10000 copies, and low abundance families ranged from 100-4000 copies. Using this estimation, 25 elements that were highly abundant and another 8 representative moderately abundant elements made up 24.5% of the haploid genome; the remaining 46 moderately abundant elements make up another 7.2%. The 24 low abundance elements make up less than 0.5% of the haploid genome. So in this estimation, the repeat families comprise 29-35% of the haploid genome of *G. hirsutum*, which roughly agrees with estimates from *Cot* analysis (Walbot and Dure 1976). Based on patterns of hybridization to genomic Southern blots, most (83/103) of the repetitive element families are interspersed or partially interspersed, with the remaining 20 being tandem or partially tandem. Based on analysis of genomic digests with isoschizomers, most interspersed repetitive elements are methylated, and most tandem repeats are not methylated (Zhao, Wing *et al.* 1995).

More recently, Hawkins *et al.* categorized repetitive sequences of *Gossypium* (Fig. 2) by BLASTing sequences from whole genome shotgun libraries against NCBI databases (Hawkins, Kim *et al.* 2006). Four genomes were randomly-sampled for sequences that resembled known repetitive element families: the A1 (*G. herbaceum*), D5 (*G. raimondii*) and K (*G. exiguum*) genomes, and an outgroup: *Gossypoides kirkii*. The repetitive sequences identified were further characterized into different groups of transposable elements and tandem repeats.

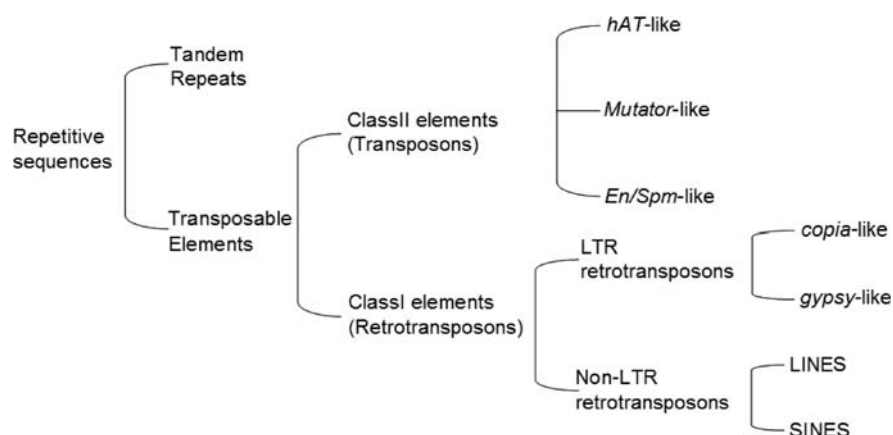


Fig. 2 The categorization of repetitive sequences in cotton

Tandem repeats were identified using *Tandem Repeat Finder* (Benson 1999). 5SrDNA was identified in all four genomes tested. The copy numbers are estimated for the D (7675 ± 3826) and A (5073 ± 3379) genomes. No significant differences were shown between the genomes, however the precision of the estimated copy numbers was relatively low (note large standard deviations). In an earlier study, the copy number of 5SrDNA was found to vary several fold even among species of the same diploid genome (Cronn, Zhao *et al.* 1996). Another previously published *Gossypium* repeat: pXP1-80 (Zhao, Si *et al.* 1998) was also identified in all four genomes, with copy numbers: *G. kirkii*: $12,263 \pm 6098$; *G. raimondii*: 6573 ± 3956 ; *G. herbaceum*: $10,101 \pm 539$ and *G. exiguum*: $23,795 \pm 8528$ (Hawkins, Kim *et al.* 2006). Some unknown types of tandem repeats were also found in low copy numbers (Hawkins, Kim *et al.* 2006).

En/Spm-like, *Mutator*-like, and *hAT*-like are the three major superfamilies of Class II (DNA) transposons identified in the cotton libraries. No evidence of MITEs, TRIMs, LARDs, or Helitrons was found in the libraries evaluated, noting however that these libraries are a relatively small sampling of the genome. *En/Spm*-like sequences make up less than 1% of the genomes of all three cotton species and the outgroup, and so does the *hAT*-like sequences. *Mutator*-like sequences were identified, but the copy numbers were not estimatable due to lack of a confident length estimate of the element. All together, Class II transposons make up only <2% of the whole genomes.

Class I transposons are much more abundant, making up about 45-60% of each of the genomes tested, indicating that these elements have amplified roughly proportionally to the size of the genome. *Copia*-like element numbers are proportional to genome sizes, except that the D genome has a higher than expected number. LINE-like elements are similar in number in the D genome and the outgroup, but have significantly higher copy numbers in the A and K genomes. SINE-like elements were not identified in these libraries. *Gypsy*-like

families have the closest relationship between copy number and relative genome size, and are considered a major component of the size differences between different cotton genomes (Hawkins, Kim *et al.* 2006).

2.3 Repetitive Sequence Evolution in Tetraploid Gossypium

The diploid origins of repetitive elements in the tetraploid cotton species can be deduced by comparative analysis with the diploid ancestors. A total of 83 non-cross-hybridizing clones from *G. barbadense* containing dispersed nuclear repetitive DNA were radioactively labeled and hybridized to quantitative slot blots of genomic DNA from a series of cotton genotypes representing the respective subgenomes (Zhao, Si *et al.* 1998). Hybridization intensities of repetitive elements are largely consistent with our present understanding of *Gossypium* phylogeny (Wendel and Albert 1992). With the exception of one D genome species (*G. gossypioides*), all the A genome “specific” elements are largely confined to closely related Old World B, E and F genomes, showing only low levels of signal in the Australian C and G genomes. The few D genome “specific” elements are confined to New World cottons, showing little signals in the Old World A, B, E and F genomes. Only 4 of the 83 repetitive fragments tested were D genome enriched or D genome specific. Most dispersed repeat families in tetraploid cotton are derived from the physically larger A-genome diploids (Zhao, Si *et al.* 1998).

The finding of otherwise A-genome specific repetitive elements in the D-genome species *G. gossypioides* is particularly interesting. The signals from A genome specific repeat probes hybridized on *G. gossypioides* are, on average, only ~36% of the level of A genome diploids, but this is 600% higher than the levels in other D genome cotton species. *G. gossypioides* is sister to *G. raimondii*, long suspected to be the closest extant relative to the tetraploid D-genome progenitor. One could envision that *G. gossypioides* may have been the tetraploid progenitor, or may have been an additional lineage spawned by the illegitimate A-D hybridization that led to polyploid formation.

The discovery of otherwise A-genome specific repetitive elements in *G. gossypioides* also suggested the possibility that repetitive elements may spread outside of their original genome following polyploid formation. Such spread has been demonstrated by fluorescence in situ hybridization (FISH): many previously A genome specific elements have spread to the D subgenome chromosomes of tetraploid cotton (Zhao, Si *et al.* 1998). The extent of spread between subgenomes varies among different families of dispersed elements: some families remain confined to the A subgenome chromosomes of tetraploid cotton (pXP137), and others (pXP224) confined to the D subgenome.

Tandemly repetitive DNA element families show evidence of concerted evolution in tetraploid cotton, with fixation of different diploid alleles in different lineages. For example, rDNA ITS sequences from 10 A, D, AD genome species and an outgroup C genome species were tested for phylogenetic

relationships. Bidirectional homogenization of tandem repeats has occurred within the AD tetraploid genome after polyploidization. One clade of tetraploid species had all rDNA homogenized to the A genome type, and the other clade had most rDNA homogenized to the D genome type (Wendel, Schnabel *et al.* 1995; Wendel, Schnabel *et al.* 1995).

3 Cotton Genome Size Evolution

Their wide range of genome sizes, well understood phylogeny, and relatively short history of divergence, makes the cotton genomes well suited to research into genome size evolution. Studies of genome size variation in cotton include two general approaches: the comparison of corresponding regions between different genomes, and global comparisons between genomes. The former approach examines closely how intron size variation, differences in the size and number of insertions and/or deletions (indels), and illegitimate recombination affect genome size. The latter compares globally the types and numbers of transposable elements between genomes.

3.1 Causes of Genome Size Variation Among Cotton Genomes

Increased genome size can be caused by polyploidization, transposable element (TE) amplification, increase in pseudogene number and/or intron size, and incorporation of organellar genome fragments into the nucleus. Polyploidization and TE amplification usually lead to large scale changes in genome size, while other mechanisms have smaller effects. Compared to the relatively established routes of genome expansion, genome size shrinkage is less well understood (Bennetzen and Kellogg 1997). However, several possible mechanisms for reduction of genome size have been suggested, including the loss of whole chromosomes, unequal intrastrand recombination, and illegitimate recombination. The loss of whole chromosomes has not yet been observed, but evidence for intrastrand recombination and illegitimate recombination has already been found in other plant genomes (Bennetzen 2002).

3.2 One Group of Class I Transposable Elements are Largely Responsible for Genome Size Variation Among Different Diploid Cotton Species

As in many plant species, genome size expansion in diploid cotton is mostly due to Class I transposable elements, i.e. retrotransposons.

Reassociation kinetics analysis showed little difference in the low copy sequences of different diploid genomes, but the complexity and copy numbers

of repetitive elements were roughly proportional to genome size (Geever, Katterman *et al.* 1989). Further, among different classes of repetitive sequences, copy numbers of tandem repeats are similar among different species. Class I transposable elements constitute 45-60% (Hawkins, Kim *et al.* 2006) of the cotton genome, and their “copy-and-paste” mechanism results in a net increase of genome size. In three different cotton genomes and an out group (A, D and K, and *Gossypioides kirkii*), the copy number of Class I elements ranged 4.4 fold: from $45,515 \pm 9241$ in the outgroup, to $197,294 \pm 18,935$ in the K genome species. The majority of repetitive elements found in these 4 genomes are LTR retrotransposons. Class II (DNA) transposable elements, using “cut-and-paste” fashion of movement, make up less than 2% of the cotton genome.

One specific group of gypsy-like retrotransposons (*Gossypium* retrotransposable gypsy-like elements group 3, i.e. Gorge3) has similar copy numbers in D genome cotton and the out group *Gossypioides kirkii* (genome size 588Mbp), but significantly higher copy numbers in the larger A and K genomes (Hawkins, Kim *et al.* 2006). From a purely quantitative standpoint, the propagation of Gorge3 family members is responsible for much variation in genome size among different *Gossypium* genomes.

3.3 Other Mechanisms of Genome Size Variation in Cotton

In addition to the effects of transposable elements, intron size differences, small indel number differences and illegitimate recombination have been examined for their possible contributions to genome size variation in cotton.

Contiguous sequence from BACs containing the cellulose synthase gene *CesA1* was compared between the two sub-genomes (At and Dt) of tetraploid cotton (*G. hirsutum*). (Grover, Kim *et al.* 2004). The overall gapped aligned length is 123.8 kb. The *CesA1* region appeared to be within a “gene island”. A total of 14 genes were detected, all present in collinear order in each of the two genomes, and totaling about 29.2 kb in size. Only two transposable elements were found, and are shared between the two homeologous genomes (Grover, Kim *et al.* 2004), indicating relatively ancient origins preceding the A-D divergence of 5-10 million years ago (Senchina, Alvarez *et al.* 2003). The high level of conservation of microsynteny in the *CesA1* region might be due to its euchromatic property. Comparative genomic research in other taxa (Bowers, Arias *et al.* 2005) has suggested that genome rearrangements may be somewhat deleterious, and more likely to happen in heterochromatic regions. It is very likely that *Gossypium* genomes, although more recently diverged from one another, may show a similar pattern, with conserved gene content and order in euchromatin and rearrangement and size variation in heterochromatin.

Integration of plastid DNA into the nucleus may contribute to cotton genome expansion. In the above study, a plastid gene, *ycf2*, inserted in the At genome, accounting for 5.6% of the At genome specific sequence of the *CesA1* region. On the other hand, intron sizes showed little difference between At and

Dt genomes (a mere gain of 3 bp in At) (Grover, Kim *et al.* 2004). Other studies concurred that there exists little intron size variation among *Gossypium* species, irrespective of genome size (Wendel, Cronn *et al.* 2002).

Small indel numbers were also evaluated in the *CesA1* BACs. Overall, small indels accounted for 14% and 18% of the total length in the At and Dt subgenome, respectively, but do not contribute significantly to the overall size difference in the region.

Among the indels discovered, 38% were flanked by short direct repeats of 2-15 bp associated with illegitimate recombination (Devos, Brown *et al.* 2002; Ma, Devos *et al.* 2004). These putative illegitimate recombinations were not equally distributed between At and Dt genomes, with the Dt genome having nearly twice as many different events as the At genome (36 vs 19); but they cover a similar amount of sequence. This suggests that illegitimate recombination is very likely a common mechanism of sequence evolution in cotton, and may also play a role in the evolution of cotton genome size.

4 Variation in the Genetic/Physical Distance Relationship

Genetic distances are measured by recombination rates between markers, but recombination events do not happen uniformly across the genome. Enormous variations in recombination frequencies exist even among different regions of the same chromosomes. This brings about variation between recombination-based genetic distances and nucleotide-based physical distances.

Genetic/physical distance variation can often be inferred based upon marker density information from genetic maps. Chromosomal regions that are densely populated with DNA markers are often characteristic of heterochromatic regions in which recombination events are rare, and therefore have a low genetic/physical distance ratio. In euchromatic regions, while there is generally more low-copy DNA (including genes) than in heterochromatic regions, this difference is outweighed by a much higher frequency of recombination, leading to an overall increase in the genetic/physical distance ratio.

To explore variations in genetic marker density, detailed cotton genetic maps composed of 2584 loci on the AD tetraploid map and 763 on the D genome map (Rong, Abbey *et al.* 2004) were used. Each linkage group was partitioned into intervals of 10 cM in length. A total of 65 intervals comprising 49 clusters were statistically marker rich. These intervals occurred in an average of 1-3 clusters on each chromosome, except for tetraploid chromosomes 1 and 25 and D-genome linkage groups D3, D6, D7, D8 and D10, with no marker-rich intervals.

On most chromosomes, at least one significant concentration of loci occurs, possibly corresponding to the centromeric regions. Virtually all marker-rich regions corresponded between the D and Dt genomes, and most also corresponded with the At genome, suggesting that these may be the locations of many of the cotton centromeres. In several cases, the breakpoints of structural

rearrangements between the A and D subgenome locate squarely in these regions (Rong, Abbey *et al.* 2004), consistent with the widespread observation that chromosomal inversion breakpoints often lay at or near centromeres. A total of three marker-rich regions are unique to Dt and nine are unique to At, generally consistent with the much larger quantity of repetitive DNA in the A genome (Zhao, Si *et al.* 1998).

A total of 17 intervals comprising 12 clusters were marker poor, all on the tetraploid genomes (Rong, Abbey *et al.* 2004). Marker-poor regions showed little correspondence, and in the At genome occurred only at the false-positive level, but did seem to be real in the Dt genome. These clues await more information about cotton genome organization to unravel their significance, if any.

5 BAC-Based Physical Mapping Projects Underway

Bacterial artificial chromosome (BAC) libraries, containing genomic DNA clones that are typically 100 kb or more in length and maintained at high fidelity by virtue of low copy-number plasmids, have proven to be valuable for study of genome organization, genome-wide physical mapping and sequencing, and isolation of key features surrounding a gene (such as promoter regions). Extensive BAC resources for global physical characterization of cotton genomes are available (Table 1). A high priority has been their use in development of scaffolds of genetically and physically-anchored sequence-tagged sites that can provide a foundation for eventual assembly of whole-genome sequences. Anchoring of these resources to DNA marker maps that have been employed in a host of genetic, evolutionary and functional studies over the past two decades, will link the eventual cotton sequences to a rich history of prior research.

A total of 10 genome-equivalent coverage of *G. raimondii* BACs have been fingerprinted at the Plant Genome Mapping Laboratory (Univ. Georgia) using standard procedures (Marra, Kucaba *et al.* 1997). To genetically-anchor the fingerprints into an integrated physical map, virtually all genetically mapped probes have been applied to the fingerprinted BACs using the overlapping oligonucleotides (overgo) method (Cai, Reneker *et al.* 1998). Manual editing and revision of the physical map is in progress, incorporating genetic marker hybridization data with BAC fingerprint data, and assembly into contigs using FingerPrinted Contigs (FPC) (Soderlund, Longden *et al.* 1997; Soderlund, Humphray *et al.* 2000). The assembly will be publicly available via a WebFPC site. Additional coverage of *EcoRI* BACs for the same genotype has recently been generated, in validation.

A BAC library of *G. hirsutum* acc. 'TM-1' has been used for whole-genome physical mapping by capillary-based technology (Xu, Sun *et al.* 2004), through collaborative research with the Kohel/Yu laboratory (USDA-ARS) and the Zhang laboratory (TAMU). A total of ~100,000 clones (~5x) have been fingerprinted on capillary sequencers. Preliminary contig assembly from the

Table 1 A summary of BAC resources known to be publicly available, and their locations

Species/ genotype	Enzyme	Insert size (kb)	Genome coverage	Source*
<i>G. hirsutum</i>				
Acala Maxxa	Hind III	137	8.3	CUGI
TM-1	Bam HI	130	4.4	TAMU
TM-1	Hind III	150	5.2	ARS
TM-1	Eco RI	175	6.0	TAMU
Auburn 623	Bam HI	140	2.7	TAMU
Tamcot HQ95	Hind III	93	2.3	TAMU
O-613-2R	Hind III	130	5.7	NAU
<i>G. barbadense</i>				
Pima S6	Hind III	100	5.0	PGML
Pima 90	Bam HI/Hind III	130	6.5	Agr Univ Hebei, China
<i>G. raimondii</i>				
unnamed acc.	Hind III	97	10.0	PGML
unnamed acc.	Eco RI	110	4.5	PGML
<i>G. arboreum</i>				
AKA8401	Mbo I	115	6.0	PGML
AKA8401	Hind III	144	9.0	PGML
<i>G. longicalyx</i>				
F1-1	Hind III	125	4.4	PGML
F1-1	Eco RI	105	3.6	PGML
<i>Gossypioides kirkii</i>				
unnamed acc.	Hind III	132	8.4	PGML
unnamed acc.	Eco RI	110	7.5	PGML

*ARS: <http://algodon.tamu.edu/cropgerm.htm>

CUGI: <http://www.genome.clemson.edu/>

NAU: cotton@njau.edu.cn

PGML: <http://www.plantgenome.uga.edu/catalog/>

TAMU: <http://hbz7.tamu.edu>

fingerprints showed that at least 20% of the resulting contigs looked to contain clones originating from homoeologous subgenomes and/or duplicated loci. To help resolve the duplicate fragments, a new TM-1 BAC library with a much larger average insert size (~175-kb) is being constructed.

Two libraries of *G. arboreum* acc. AKA8401, totaling about 15 genome-equivalent coverage, are being genetically-anchored by hybridization to genetically-mapped DNA probes. These data will be incorporated into the existing 'BACMan resource' at the Plant Genome Mapping Laboratory web site (www.plantgenome.uga.edu), which already includes similar anchoring data for BAC libraries for *G. hirsutum* 'Acala Maxxa', *G. barbadense* 'Pima S6', and *G. raimondii*.

A BAC library from a male-sterile fertility restorer line 0-613-2R (*G. hirsutum* L.) has been used for identification of Rf1 gene in a 100-kb region (Yin, Guo *et al.* 2006). FISH of landed BACs recently completed the assignment of linkage groups to identified chromosomes (Wang, Song *et al.* 2006).

6 Perspectives

6.1 *Implications of Physical Organization of the Gossypium Genomes for Whole-Genome Sequencing*

Efficient strategies for capturing the sequence diversity represented within the *Gossypium* genus will be greatly influenced by the large differences in genome size and organization that differentiate species and genome types within the genus. The 3-fold variation in diploid genome size appears to have accumulated in about 5-10 million years since the diploid clades diverged from a common ancestor (Senchina, Alvarez *et al.* 2003). Much of this genome size variation is due to dispersed repetitive DNA (Zhao, Si *et al.* 1998), which appears to be largely LTR retrotransposon-like elements (Hawkins, Kim *et al.* 2006). There have been particularly large expansions of repetitive DNA content in the A/B/E/F and C/G/K clades in the 5-10 million years since their divergence, thus many repetitive element families in these clades may include large numbers of relatively recently-derived members – this condition would be especially problematic for whole-genome shotgun sequencing approaches, which require individual sequencing reads to be distinguishable (even if only by a single nucleotide) from all other sequences in the genome. By contrast, the D genome clade appears to have few such recently-amplified repetitive DNA families, and is expected to be more amenable to whole-genome shotgun approaches, that permit rapid production and assembly of a sequence with a minimum of background information (although favored by, and fully incorporating, any such information that exists, such as genetic and physical maps). The fact that there exists a high degree of colinearity and synteny among the A, D, and tetraploid genomes (Reinisch, Dong *et al.* 1994; Brubaker, Paterson *et al.* 1999; Rong, Abbey *et al.* 2004; Desai, Chee *et al.* 2006) suggests that complete sequencing of a D-genome genotype by an economical whole-genome shotgun approach, together with reduced-representation sequencing of representatives of additional branches of the *Gossypium* family tree by a combination of EST sequencing, Cot-based, and methylation-based methods, might be a cost-effective means to quickly capture much of the genomic diversity among the diploid cottons.

DNA content of the allopolyploids is approximately the sum of those of the A and D-genome progenitors. However, recent polyploidy introduces new dimensions into the evolution of these genomes. The tetraploid clades combine the properties of the A and D genome diploids with modification by intergenomic concerted evolution, already clearly documented for the repetitive DNA fraction (Wendel, Schnabel *et al.* 1995; Wendel, Schnabel *et al.* 1995; Cronn, Zhao *et al.* 1996; Zhao, Si *et al.* 1998). The possibility of intergenomic exchange of low-copy DNA remains somewhat unclear, with tenuous evidence for it from genetic mapping (Reinisch, Dong *et al.* 1994), and against it from localized comparisons of small numbers of corresponding sequences (Cronn, Small *et al.* 1999), but growing data from other taxa strongly suggest that it may be an important

dimension of polyploid evolution (Hughes and Hughes 1993; Moore and Purugganan 2003; Gao and Innan 2004; Chapman, Bowers *et al.* 2006). Recent data from computational analysis of the rice genome suggests concerted evolution of even low-copy sequences that are diverged by a few million years (Wang, Tang *et al.* Accepted), roughly the degree of divergence among the cotton diploids. In the tetraploid cotton genome(s), the possibility of intergenomic concerted evolution both among repetitive and low-copy DNA families may strengthen the case for a BAC-based rather than a whole-genome shotgun approach.

6.2 Future Directions

Despite much progress (detailed above), there still exist numerous gaps in infrastructure and information needed to clarify our knowledge of cotton genome structure.

First and foremost, the *Gossypium* community lacks a high-quality reasonably complete genome sequence to use as a reference, the nearest one phylogenetically being that of *Arabidopsis* (Rong *et al.* 2007) and of some value but also suffering numerous limitations. A recent investment by the US Department of Energy Joint Genome Institute 'Community Sequencing Program' will provide about 0.5 genome-equivalent coverage of *G. raimondii*, sufficient to clarify whether this smallest and least repetitive of *Gossypium* genomes is amenable to whole-genome shotgun sequencing, guided by its genetic (Rong *et al.* 2004) and physical (see above) maps.

Second, we need not only sequence one diploid progenitor, but both, and a tetraploid *Gossypium* sequence as well. A host of data show that the polyploid formation and associated ~1-2 million year period of adaptation to the polyploid state, have been of both fundamental and practical importance in *Gossypium* evolution and improvement. Issues raised above regarding the degree of homogeneity of repetitive fractions, and the degree of intergenomic concerted evolution of low-copy DNA that has taken place, need to be clarified in order to formulate an effective strategy for this undertaking. The *Gossypium* community is acutely aware of these needs, and actively working to bring them to fruition.

Third, a host of interesting and potentially important genetic variation exists within members of the *Gossypium* genus that are difficult to access by sexual crosses. Further progress is needed to complete BAC resources for the various genome types (and preferably for multiple diverse representatives within each type), and to use multiple complementary approaches detailed above to extend *Gossypium* sequence information to these additional taxa.

References

- Bennetzen, J. L. (2002). "Mechanisms and rates of genome expansion and contraction in flowering plants." *Genetica* **115**(1): 29–36.
- Bennetzen, J. L. and E. A. Kellogg (1997). "Do Plants Have a One-Way Ticket to Genomic Obesity?—*Plant Cell* **9**(9): 1509–1514.

- Benson, G. (1999). "Tandem repeats finder: a program to analyze DNA sequences." *Nucleic Acids Res* **27**(2): 573–80.
- Bowers, J. E., M. A. Arias, et al. (2005). "Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses." *Proc Natl Acad Sci U S A* **102**(37): 13206–11.
- Brubaker, C. L., A. H. Paterson, et al. (1999). "Comparative genetic mapping of allotetraploid cotton and its diploid progenitors." *Genome* **42**: 184–203.
- Cai, W. W., J. Reneker, et al. (1998). "An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization." *Genomics* **54**(3): 387–97.
- Chapman, B. A., J. E. Bowers, et al. (2006). "Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication." *Proc Natl Acad Sci U S A* **103**(8): 2730–5.
- Cronn, R. C., R. L. Small, et al. (1999). "Duplicated genes evolve independently after polyploid formation in cotton." *Proc Natl Acad Sci U S A* **96**(25): 14406–11.
- Cronn, R. C., X. Zhao, et al. (1996). "Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons." *J Mol Evol* **42**(6): 685–705.
- Desai, A., P. W. Chee, et al. (2006). "Chromosome structural changes in diploid and tetraploid A genomes of *Gossypium*." *Genome* **49**(4): 336–45.
- Devos, K. M., J. K. Brown, et al. (2002). "Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*." *Genome Res* **12**(7): 1075–9.
- Gao, L. Z. and H. Innan (2004). "Very low gene duplication rate in the yeast genome." *Science* **306**(5700): 1367–70.
- Geever, R. F., F. R. H. Katterman, et al. (1989). "DNA hybridization analyses of a *Gossypium* allotetraploid and two closely related diploid species." *Theor Appl Genet* **77**: 553–559.
- Grover, C. E., H. Kim, et al. (2004). "Incongruent patterns of local and global genome size evolution in cotton." *Genome Res* **14**(8): 1474–82.
- Hawkins, J. S., H. Kim, et al. (2006). "Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*." *Genome Res* **16**(10): 1252–61.
- Hughes, M. K. and A. L. Hughes (1993). "Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*." *Mol Biol Evol* **10**(6): 1360–9.
- Kadir, Z. (1976). "DNA evolution in the genus *gossypium*." *Chromosoma* **56**: 85.
- Ma, J., K. M. Devos, et al. (2004). "Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice." *Genome Res* **14**(5): 860–9.
- Marra, M. A., T. A. Kucaba, et al. (1997). "High throughput fingerprint analysis of large-insert clones." *Genome Res* **7**(11): 1072–84.
- Moore, R. C. and M. D. Purugganan (2003). "The early stages of duplicate gene evolution." *Proc Natl Acad Sci U S A* **100**(26): 15682–7.
- Paterson, A. H. (2006). "Leafing through the genomes of our major crop plants: strategies for capturing unique information." *Nat Rev Genet* **7**(3): 174–84.
- Peterson, D. G., S. R. Schulze, et al. (2002). "Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery." *Genome Res* **12**(5): 795–807.
- Peterson, D. G., S. R. Wessler, et al. (2002). "Efficient capture of unique sequences from eukaryotic genomes." *Trends Genet* **18**(11): 547–50.
- Reinisch, A. J., J. M. Dong, et al. (1994). "A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome." *Genetics* **138**(3): 829–47.
- Rong, J., C. Abbey, et al. (2004). "A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*)." *Genetics* **166**(1): 389–417.

- Senchina, D. S., I. Alvarez, et al. (2003). "Rate variation among nuclear genes and the age of polyploidy in *Gossypium*." *Mol Biol Evol* **20**(4): 633–43.
- Soderlund, C., S. Humphray, et al. (2000). "Contigs built with fingerprints, markers, and FPC V4.7." *Genome Res* **10**(11): 1772–87.
- Soderlund, C., I. Longden, et al. (1997). "FPC: a system for building contigs from restriction fingerprinted clones." *Comput Appl Biosci* **13**(5): 523–35.
- Walbot, V. and L. S. Dure, 3rd (1976). "Developmental biochemistry of cotton seed embryogenesis and germination. VII. Characterization of the cotton genome." *J Mol Biol* **101**(4): 503–36.
- Walbot, V. and L. S. Dure, 3rd (1976). "The developmental biochemistry of cotton seed embryogenesis and germination. VII. Characterization of the cotton genome." *Biochim Biophys Acta* **101**: 503–536.
- Wang, K., X. Song, et al. (2006). "Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping." *Theor Appl Genet* **113**(1): 73–80.
- Wang, X., H. Tang, et al. (Accepted). "Extensive concerted evolution of rice paralogs and the road to regaining independence." *Genetics*.
- Wendel, J. F. and V. A. Albert (1992). "Phylogenetics of the Cotton genus (*Gossypium*): Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications." *Systematic Botany* **17**(1): 115–143.
- Wendel, J. F., R. C. Cronn, et al. (2002). "Intron size and genome size in plants." *Mol Biol Evol* **19**(12): 2346–52.
- Wendel, J. F., R. C. Cronn, et al. (2002). "Feast and famine in plant genomes." *Genetica* **115**(1): 37–47.
- Wendel, J. F., A. Schnabel, et al. (1995). "Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*)." *Proc Natl Acad Sci U S A* **92**(1): 280–4.
- Wendel, J. F., A. Schnabel, et al. (1995). "An unusual ribosomal DNA sequence from *Gossypium gossypoides* reveals ancient, cryptic, intergenomic introgression." *Mol Phylogenet Evol* **4**(3): 298–313.
- Xu, Z., S. Sun, et al. (2004). "Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality." *Genomics* **84**(6): 941–51.
- Yin, J., W. Guo, et al. (2006). "Physical mapping of the Rf1 fertility-restoring gene to a 100 kb region in cotton." *Theor Appl Genet* **112**(7): 1318–25.
- Zhao, X., R. A. Wing, et al. (1995). "Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L.)." *Genome* **38**(6): 1177–88.
- Zhao, X. P., Y. Si, et al. (1998). "Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton." *Genome Res* **8**(5): 479–92.

The *Gossypium* Transcriptome

Joshua A. Udall

Abstract cDNA sequences representing the *Gossypium* transcriptome have been generated by numerous researchers within the cotton community. The bulk of these cDNA sequences are from cultivated *Gossypium hirsutum* and its close diploid relatives *G. arboreum* and *G. raimondii*. Our current understanding of the *Gossypium* transcriptome is based on the assembly of these expressed sequences into unique, non-overlapping unigenes (putative gene models). With this compilation of expressed sequences, DNA microarrays have been constructed for gene expression experiments, evidence for recent genome duplication has been identified, and novel expression patterns in fiber development have been discovered. Future efforts to collect expressed sequences from domesticated cotton and its relatives will continue to enhance our understanding of the *Gossypium* transcriptome.

1 Introduction

The *Gossypium* transcriptome refers to all transcribed sequences in any of approximately 45 diploid and 5 tetraploid species in the genus. Four different *Gossypium* species have been domesticated for fiber production and genes expressed in cotton fiber have been extensively studied because of their economic importance. Other aspects of the *Gossypium* transcriptome are only beginning to receive wide-spread attention. Consequently, our current knowledge of the *Gossypium* transcriptome as a whole is heavily biased toward genes expressed in fibers.

Qualitative or quantitative differences in *Gossypium* fiber morphology represent genetic differences in coding regions or in gene regulation. Qualitative, heritable differences have been identified through complementary crosses of fiber mutants (Kohel et al. 2002). Quantitative, heritable differences in fiber

J.A. Udall (✉)

Plant and Wildlife Science Dept, Brigham Young University
e-mail: jaudall@byu.edu

quality have been identified through inter-specific quantitative trait loci (QTL) mapping studies (Wright et al. 1998; Jiang et al. 1998; Paterson et al. 2003; Lacape et al. 2005). Obtaining gene sequences responsible for this heritable variation, measuring their respective expression levels, and associating genetic differences with phenotypic traits is a central focus of cotton genetic research. Two of these three foci, obtaining gene sequences and measuring their respective expression levels, directly characterize the *Gossypium* transcriptome. Herein, I summarize progress toward characterization of the *Gossypium* transcriptome, through direct sequencing of expressed sequence tags (ESTs), EST analysis *in silico*, and EST-derived microarrays.

2 ESTs

High-throughput sequencing has provided significant insights into the transcriptomes of numerous plant species including cotton. Transcriptomes can be surveyed by cloning cDNA generated from poly-A-containing mRNA transcripts, although a significant portion of cDNA transcripts represent only partial gene sequences. cDNA can be selected or enriched for genes expressed in response to a particular environmental treatment or from a particular tissue type by using subtractive hybridization (Straus and Ausubel 1990) or differential display techniques (Liang and Pardee 1992). Alternatively, sequencing many random clones from cDNA libraries can also provide a significant amount of gene sequence information. In both cases, partial or full gene sequence derived from an expressed gene is known as an Expressed Sequence Tag (EST), where the word 'tag' implies that the entire gene sequence may not have been sequenced. EST sequencing projects have been completed or are underway for cotton (Arpat et al. 2004; Wu et al. 2005; Yang et al. 2006; Amarasinghe et al. 2006; Udall et al. 2006; Taliercio and Boykin 2007). In other plant species, EST sequences have provided novel insights into genome structure and redundancy, (Schlueter et al. 2004), inter-genomic comparisons (Fulton et al. 2002), and gene discovery (Ewing et al. 1999; Ronning et al. 2003; Hughes and Friedman 2004). Similar insights have been made in cotton using recently derived ESTs.

2.1 EST Collection and Contributors

Several institutions and laboratories have contributed to the public (GenBank) collection of *Gossypium* ESTs (Table 1). Four large EST contributions (Arpat et al. 2004; Yang et al. 2006; Udall et al. 2006; Taliercio, unpublished) comprise 72% of the ESTs in Genbank and were funded by either the National Science Foundation Plant Genome Research program or the USDA-ARS. Smaller, but significant, contributions have been made from numerous other cotton genome research projects and represent transcriptional profiles from either specialized tissues or specific experimental treatments (Table 1).

Table 1 Summary of *Gossypium* GenBank EST libraries used for assembly construction. Sequencing (S) was in one or both directions (f = 5'; r = 3'; x or y = orientation unspecified). Quality values (Q) are PHRED scores (Ewing et al. 1998) calculated from the shape and area of fluorescence intensity from each base pair when sequenced on a fluorescent automated sequencer. The library acronym has a two letter prefix indicating species (GH = *G. hirsutum*, GR = *G. raimondii*, and GA = *G. arboreum*)

Species	Authors	Accession	Library	Library description	# ESTs	Q
<i>G. arboreum</i> (A ₂)	Wing et al., Arpat et al.	AKA8401	GA_Ea	7-10 dpa fibers (normalized)	46,603f + r	y
<i>G. raimondii</i> (D ₅)	Udall et al.	GN34	GR_Ea	Whole seedlings with first true leaves	33,671f + r	y
<i>G. hirsutum</i> (AD ₁)	"	"	GR_Eb	-3 dpa buds to + 3 dpa bolls	35,061f + r	y
	Allen	Coker 312	GH_MDI	8-10 dpa boll (irrigated)	1,144f	y
	"	"	GH_MDDS	8-10 dpa boll (drought stressed)	1,238f	y
	Allen & Payton	"	GH_LDI	15-20 dpa boll (irrigated)	1,799f	y
	"	"	GH_LDDS	15-20 dpa boll (drought stressed)	1,409f	y
	Blewitt & Burr	Acala Maxxa	GH_BNL	Fiber 5 dpa (normalized)	8,022x	n
	Chapman	Stw 7A gl	GH_ECT	18 hr etiolated seedlings	2,880x	y
	Dowd & McFadden	Delta	GH_CRH	Root and hypocotyls	1,464f	y
	"	Emerald	GH_CFUS	RH tissues, <i>Fusarium oxysporum</i> infected	820f + x	y
	Faivre-Nitschke & Dennis	Sicot	GH_LSL	Sicot S9; leaves, late season	1,810f	y
	Grou & Chen	Xu-142	GH_FOX	Ovule (0-5dpa) and fiber (1-22 dpa)	7,997f	y
	Haigler & Wilkerson	Delta Pine 90	GH_SCW	Secondary vs. primary fiber (subtr.)	7,385x	y
	Klueva et al.	Coker 312	GH_SDL	Seedling (control)	1,918f	y
	Klueva & Nguyen	"	GH_SDL	Seedling (drought stressed)	1,142f	y
	"	"	GH_SDC	Seedling (chilled)	576f	y
	Liu & Dennis	Delta Pine 16	GH_IME	Immature embryo	1,536x	y
	Patil, Essenberg, Pierce	Im216	GH_IMX	Leaf 8, 14, 20, 30, 45, 60 hpi <i>Xanthomonas</i>	1,134x	y
	"	AcB4b7In	GH_ACXE	Leaf 8 + 14 hpi <i>Xanthomonas</i>	647x	y

(continued)

Table 1 (continued)

Species	Authors	Accession	Library	Library description	# ESTs	Q
"	"	"	GH_ACXM	Leaf 20 + 30 hpi <i>Xanthomonas</i>	1,328x	y
"	"	"	GH_ACXL	Leaf 45 + 60 hpi <i>Xanthomonas</i>	862x	y
Suo & Xue		Zhongmian12	GH_SUO	0 dpa ovule	1,240x	n
Trolinder		T25	GH_pAR	Leaves	1,230x + y	y
Taliercio		DES119	GH_STEM	Mature stem	8,643x + y	y
Ni, Trelease		DP62	GH_ECOT	Etiolated cotyledon	2,772x	y
Wan & Wing		91-D-92	GH_CBAZ	Cotton boll abscission zone cDNA Library	1,306f	y
Wu & Dennis		Delta Pine 16	GH_CHX	Ovules -3 to 0 dpa cycloheximide	7,631x + y	y
"	"	"	GH_OCF	Ovules 0 dpa	867f	y
"	"	"	GH_ON	Ovules 0 dpa normalized	5,903 _f ^{f+r}	y
Taliercio, Udall, Wen-del, Scheffler		DES119	GH_FFF	Fiber 5-dpa	1,603f + r	y
"	"	"	GH_FFT	Fiber 3-dpa	274f + r	y
"	"	"	GH_HYPS	Stem (lowest three nodes), 7 wk old plts	1,029f + r	y
"	"	"	GH_HYPT	Stem, 3 wk old plants	3,203f + r	y
"	"	"	GH_ONEFM	Fiber 1-dpa (RNA from membrane-bound polyribosomes)	1,641f + r	y
"	"	"	GH_OVFIB	Ovules 0-dpa	1,749f + r	y
"	"	"	GH_OVfib	Ovules 0-dpa (fiberless mutant)	1,727f + r	y
"	"	"	GH_RMIRS	Mixed ages of roots, stems, meristems, and 1-5d fiber initials	19,600f + r	y
"	"	"	GH_RTFR	Root, 7-10 wk plants (free ribosome)	2,081f + r	y
"	"	"	GH_RTMB	Root, 3 wk plants (membrane-bound ribosome)	669f + r	y

(continued)

Table 1 (continued)

Species	Authors	Accession	Library	Library description	# ESTs	S	Q
"	"	"	GH_RTMTF	Root, 3 wk plants (unbound ribosomes)	2,818f	+ r	y
"	"	"	GH_STEMmb	Stem, 7-10 wk plants (membrane-bound polyribosomes)	519f	+ r	y
"	"	"	GH_TMIRS	Mixed ages, roots, stems, meristems, 1-5 dpa fiber initials	25,015f	+ r	y
"	"	"	GH_TS	Stem, 3-5 wk plants	3,087f	+ r	y
"	"	"	GH_TSMB	Stem, 3-5 wk plants (membrane bound)	690f	+ r	y
"	"	"	GH_TSnorm	Stem, 3-5 wk plants, normalized	1036f	+ r	y
"	"	"	GH_YF	Fiber, 1-3 dpa	1,583f	+ r	y
Yang et al.	TM-1		GH_TMO	Ovules, -3 to + 3 dpa	32,190f	+ r	y
Shi et al.	Xuzhou 142		GH_FIB1	5-10 dpa fiber	11,692f		y
Total # ESTs					285,726		

¹ ESTs of the GH_FIB library were assembled prior to their submission to GenBank.
² Small differences in the number of ESTs between the totals listed here and totals found in GenBank are due to low-complexity removal (or not) before or after GenBank submission

As listed in Table 1, *Gossypium* ESTs have been derived from three different species of cotton: *Gossypium arboreum* (A₁-genome), *G. hirsutum* (AD₁-genome), and *G. raimondii* (D₅-genome). The phylogenetic relationships between cotton genome groups and their member species have been thoroughly studied (Cronn et al. 2002; Senchina et al. 2003). Diploid *G. arboreum* and *G. herbaceum* (A₂ and A₁, respectively) are phylogenetically equidistant to the A_T genome of allopolyploid cotton (Cronn et al. 2002; Wendel and Cronn 2003). Diploid *G. raimondii* is the D-genome species most closely related to the modern-day allopolyploid D_T genome (Endrizzi et al. 1985; Wendel 2000; Wendel and Cronn 2003). Genetic divergence between diploid groups (A vs D) has been estimated to be 0.009 ± 0.014 (K_a) and 0.038 ± 0.036 (K_s). Genetic divergence between the separate genomes of the allopolyploid (A_T – D_T) have been estimated to be 0.010 ± 0.013 (K_a) and 0.042 ± 0.025 (K_s), where the ‘T’ subscript indicates that these genomes are in a tetraploid nucleus (Senchina et al. 2003). Similar divergence estimates between the diploid and tetraploid sequences suggest a recent origin (approximately 1.3 MYA assuming 2.6×10^{-9} substitutions/site/year; Gaut 1998). Because of these low sequence divergence rates between the A and D genomes (diploid or polyploidy), individual EST sequences from cDNA libraries have been combined into a single EST assembly.

2.2 EST Assembly

A collection of EST sequences represents a random sample of mRNA transcripts, with genes expressed at high levels more frequently represented than genes expressed at low levels. Various informatic and laboratory techniques can be used to reduce the copy number of the highly expressed transcripts (Straus and Ausubel 1990; Liang and Pardee 1992), but no method completely eliminates multiple ESTs from the same transcript. Repeated sampling of highly expressed transcripts can be overcome by bioinformatic techniques that assemble the overlapping segments of repeated sequences into a single consensus sequence or contig (Ewing and Green 1998; Huang and Madan 1999; Fig. 1). The consensus sequence represents a putative gene sequence. ESTs of genes expressed at low-levels will often not have sequence overlap with any other EST in the dataset, being represented as ‘singletons’. The total of consensus plus singleton sequences define a set of *unigene* or tentative consensus (TC) sequences.

Cotton EST assemblies have been both created and viewed by two separate programs: the Program for Annotating and Viewing ESTs (PAVE) developed at the University of Arizona (<http://www.agcol.arizona.edu/>), and the Cotton Gene Index at the Computational Biology and Functional Genomics Laboratory located at the Dana-Farber Cancer Institute and the Harvard School of Public Health (CBFGL; formerly part of The Institute of Genomic Research).

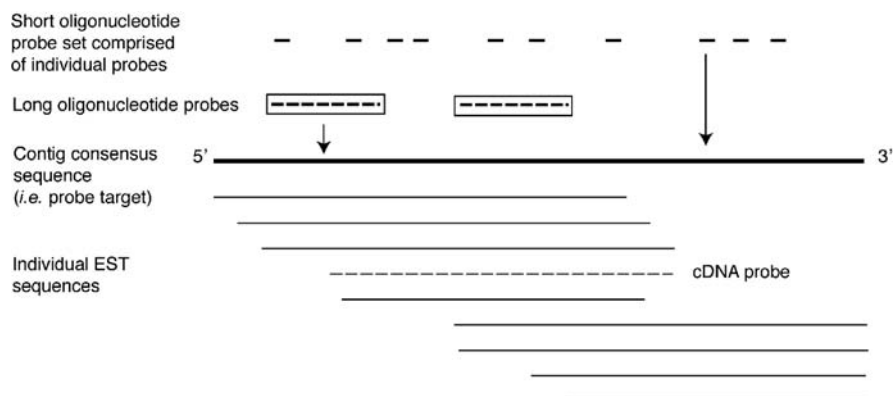


Fig. 1 An assembled EST contig targeted by multiple microarray probe types. Two arrows exemplify hybridization that occurs between each probe and cDNA target on a DNA microarray

Both programs operate in similar fashion. The quality of the EST assembly largely depends on high quality EST sequence and several assembly parameters such as a percent identity threshold that determines which ESTs are assembled into a single contig.

Comparisons of the distinct unigene sets suggest that the similarity between the two assembly programs is high, but some homologous contig alignments generated by these programs may be different. These differences are a result of the internal algorithms of each program. At the core of PAVE, Phrap assembles ESTs with some added functional constraints (Ewing et al. 1998). PAVE runs a self-blast with all the ESTs and creates broad clusters (single linked-clusters) of ESTs based on sequence homology. These clusters are then independently assembled into contigs based on sequence homology. A feature of PAVE does not allow mate-pairs (forward and reverse sequences of a single clone) to be split across contigs. If mate-pairs do not overlap, they are joined by n's. It also assigns the mate-pairs to the same contig if each sequence identifies the same protein through BLASTX to swiss-prot. This functions as a safe-guard in case a clone is chimeric or the ESTs are mis-named. Collectively, the PAVE contig consensus sequences and singletons (referred to as unigenes) are blasted against UniProt.

The core of CBFGL is a proprietary Paracel Transcript Assembler. First, EST sequences for cotton are extracted from the National Center for Biotechnology Information's (NCBI's) dbEST when a gene index is being built (*i.e.* EST assembly). During each assembly iteration, the 'new' EST sequences are screened for quality control purposes (vector, E.coli, poly-A and poly-T trimming, length ≥ 100 bp, $<3\%$ ambiguous bases *i.e.* N). ESTs are compared and clustered together if they meet the following criteria: a minimum of 40 base pair match, greater than 94% identity in the overlap region, and a

maximum unmatched overhang of 30 base pairs. With these parameters, the transcript assembler uses a similar approach as PAVE of clustering and assembling the individual ESTs.

Both of these programs are important to the *Gossypium* transcriptome because their respective EST assemblies can be browsed or searched by the cotton community. The dynamic nature of EST information (*i.e.* sequences continue to accumulate) necessitates browse and search capabilities of EST assemblies for current and future genomic research. Both PAVE and the CBFGL assemblies have a CGI-based web interface to an EST database that allows the users to search for individual ESTs or browse through individual contigs. Using the built in search functions, one can identify contigs that have ESTs from specified libraries but no ESTs from other libraries. For EST IDs, PAVE uses community derived names that identify the species and cDNA library of each EST (Table 1) while CBFGL uses the GenBank ID as the EST ID.

A final methodology was used to create the most comprehensive EST assembly by ESTInformatics (<http://www.estinformatics.org>). Currently, this platform does not have web-based browsing, viewing, or EST search capability. The ESTInformatics assembler resembles the TIGR Gene Index Cluster (TGICL) assembly tool (Perte et al. 2003) where sequences are assigned to clusters based on sequence similarity, and then contigs are assembled as individual clusters of ESTs. However, the ESTInformatics assembler contains some custom modifications of the TGICL tool further automating the assembly pipeline.

Between assemblies, the total number of cotton unigenes has loosely depended on the number of available EST sequences and the parameters used for the assembly process (Fig. 2). Six distinct, global assemblies have been

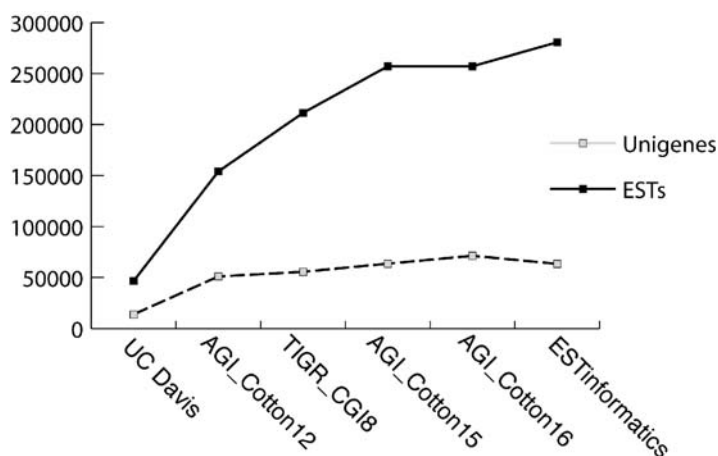


Fig. 2 The number of ESTs and number of unigenes in the cotton transcriptome across multiple EST assemblies

reported in functional genomics applications (Wilkins and Arpat 2005; Udall et al. 2006; Udall et al. 2007; <http://ESTinformatics.org>; Flagel, Udall, Nettleton, and Wendel, unpublished). To an extent, the number of unigenes have increased as the number of individual EST sequences increased (Fig. 2). Contributions of ~68,000 ESTs in 2006 by the USDA-ARS (Talierecio, personal communication) represented a significant increase in the number of EST, yet the number of unigenes appears to have decreased in the ESTinformatics assembly. While this small decrease could simply be due to differences in assembly parameters, it is striking how closely this total number of unigenes (54,085) agrees with recent estimates of diploid cotton gene number of 53,550 (Rabinowicz et al. 2005) based on methyl-filtration estimates and $50,128 \pm 7,489$ based on sequencing whole genome shotgun libraries (J. Hawkins, J. Nason, and J.F. Wendel, personal communication).

The general transcriptome of cotton (ESTinformatics assembly) is comparable to that of other angiosperms. Sixty-four percent of cotton unigenes identified a homologous protein in the NCBI non-redundant protein sequence database (BLASTX, e-value threshold $1e^{-15}$; Fig. 3). Of these blast reports, the top hit was based on homologies between cotton unigenes and several different plant species including *Arabidopsis thaliana* (54%), *Medicago truncatula* (13%), and *Oryza sativa* (10%). The only non-plant organism identified by 10 or more *Gossypium* ESTs was a *Fusarium* fungus (189 *Gossypium* ESTs best hits). These fungal sequences were likely cloned as part of the GH_CFUS library (Dowd et al. 2004). It is not surprising that the majority of unigenes had their highest homology to *Arabidopsis* given that *Arabidopsis* is cotton's closest relative with copious genomic information (Blanc and Wolfe 2004). Genes whose top blast hit was from other plant species may 1) represent genes that were lost in *Arabidopsis* (due to genome size reduction) but are common in other higher plants (e.g. *Gossypium* and *Medicago*) or 2) may represent genes that evolved more quickly in *Arabidopsis* than in other species. In the former case, it is unlikely that many genes present in *Gossypium* were lost in *Arabidopsis*. Indeed, only 6% of the genes had homology to other plants but not to *Arabidopsis*. Regarding the latter case, many gene homologies were found in separate comparisons of the cotton transcriptome to curated protein databases such as *Arabidopsis* (98% of nr total), rice (88% of nr total), and poplar (100% of nr total) suggesting that a rather large set of genes are conserved among several plant species and that blast alignment scores of individual cotton genes either occasionally misrepresent phylogenetic relationships or are different for individual genes.

Because of a high degree of coding sequence homology between cotton and *Arabidopsis*, much of the functional genomic information of *Arabidopsis* can be extrapolated to cotton (see Comparative Genomics of Cotton and *Arabidopsis* (Rong), this volume). Predicted coding sequence and putative gene function of *Arabidopsis* are routinely assigned non-exclusive, descriptive terms of molecular function, biological component, or cellular component known as gene ontology (GO). The GO terms of *Arabidopsis* genes can be used to annotate putative cotton ESTs once gene-to-gene relationships have been made with BLASTX.

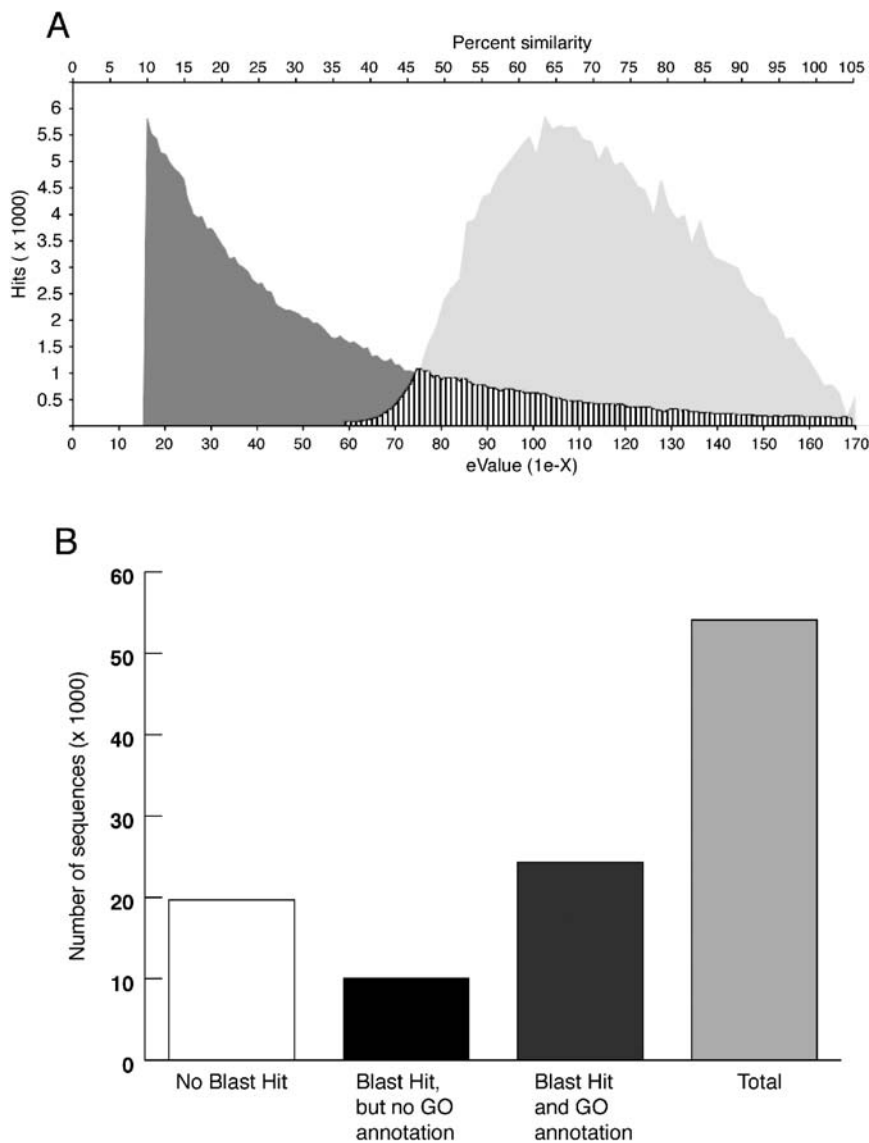


Fig. 3 Blast results of the ESTInformatics assembly when queried against *Arabidopsis* proteins (TAIR7_2007425). A) The number of sequences having a hit to an *Arabidopsis* protein at each BLASTX alignment score lower than $1e^{-15}$ and percent similarity. B) The distribution of cotton unigenes with a BLASTX hit and the number of hits that contained a GO annotation

Many more software tools exist for ‘extrapolating’ GO terms than can be mentioned here; however, Blast2GO has a user-friendly interface and has been used to assign putative gene function of cotton unigenes via GO terms (Conesa et al. 2005).

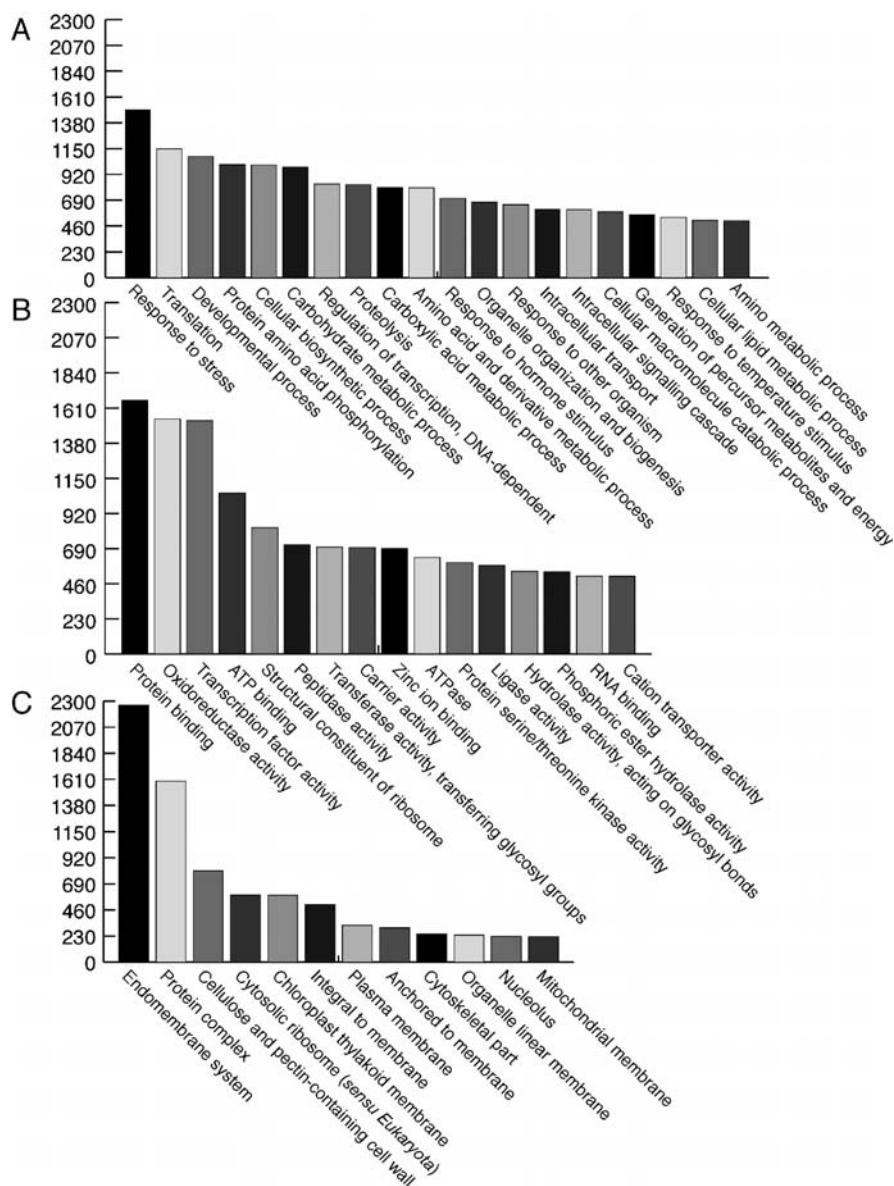


Fig. 4 Frequency of gene ontology terms associated with cotton unigenes. a) Biological process, b) Molecular function, c) Cellular components

Using a threshold of $1e^{-15}$, 94% of the cotton unigenes with an *Arabidopsis* BLASTX hit were assigned a GO term (Fig. 4). These terms represent a broad sample of gene functions, particularly because of the numerous and varied cDNA libraries that contributed to the public EST effort. Many of these

putative gene functions may be of particular interest to cotton researchers such as the molecular function of transcription factors ($n = 1528$) that may induce the developmental changes of cotton fiber from an ovule outer integument cell to mature fibers, the biological process of stress response ($n = 1,495$) that may regulate the amount of water entering the vacuole of developing fiber, or cellular components of the endomembrane system ($n = 2,264$) that may provide cellular components needed for cell growth.

3 A Transcriptome of Ancient and Recent Duplicated Genes

3.1 Ancient Duplications

Expression of duplicated genes contributes to the transcriptome of *Gossypium* and their level of expression may be impacted by polyploidy. Polyploidy is a term used to describe genomes that have undergone a whole genome duplication event. As a result of polyploidy, every gene in the genome is instantaneously duplicated and duplicate copies are referred to as homoeologs. For genes under purifying natural selection, each homoeolog shares its selection pressure with its 'copy' because of its functional redundancy. Thus, ancient genome duplication events may have reduced purifying selection pressure on individual gene copies and allowed for divergence both in terms of expression (see Li et al. 2005 for recent review) and protein function (Wilkins and Jernstedt 1999; Li et al. 2005). Evidence of ancient polyploidization events (paleopolyploidy) that occurred millions of years ago can be found in diploid genomes.

Two methods have been used to identify duplicated genome structure signifying an ancient duplication. One method is through map-based alignments (see Comparative Genomics of Cotton and *Arabidopsis* (Rong), this volume). A second method is through EST-mining (Blanc and Wolfe 2004; Schlueter et al. 2004; Cui et al. 2006). Duplicated genes within EST unigene sets can be identified by a self-BLASTX analysis. Each unigene will hit itself or an unassembled EST as the top hit, but the next best hit is potentially an ancient paleolog. After identification of paralogs or putative paleo-homoeologs, DNA and protein alignments are constructed to estimate synonymous (K_S) and non-synonymous substitution (K_A) rates. Plotting the K_S values (number of substitutions per synonymous site) against the number of gene pairs results in a distribution of points containing (or not containing) peaks representing ancient duplication events (Fig. 5). Thus, diploid cotton species appear to have undergone at least two rounds of ancient genome duplication and these peaks may correspond to the alpha and beta paleo-polyploidy events shared with other angiosperms (Bowers et al. 2003). As one might imagine, this approach is sensitive to quality of EST assembly and refinements to this approach have appeared with each subsequent publication that uses a

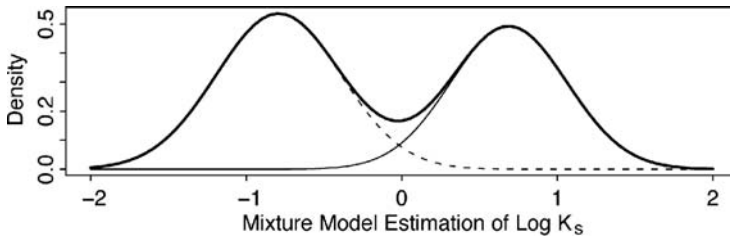


Fig. 5 Putative distribution of paleo-homoeologs in *G. arboreum* where density is similar to a histogram but the dataset is modeled as a continuous distribution rather than discrete bins. The $\text{Log } K_s$ values on the x-axis can be interpreted as ‘time’ and the density on the y-axis can be interpreted as number of duplicated gene pairs

BLAST approach (Schlueter et al. 2004; Blanc and Wolfe 2004; Cui et al. 2006). Some additional refinements have been used to analyze the ESTs of *Gossypium* including 1) skipping the perfect matches in a self-BLAST output (these are probably due to unassembled sequences) 2) using the top several blast hits instead of just the top hit [using only the top blast hit predisposes recovery of a single peak where multiple peaks may exist (*i.e.* alpha and beta)], and 3) placing paleo-duplicated sequences in a phylogenetic context including *Gossypium*’s relationship to *Arabidopsis* and *Theobroma*. In independent analyses of *G. arboreum*, *G. raimondii*, and *G. hirsutum* ESTs, approximately 4,800 pairs of duplicated genes were found in the primary and secondary duplication events (excluding the current polyploids; Fig 5). When comparing the sequences of these duplicated genes to genes from another closely related genus, *Theobroma*, it appeared that *Gossypium* and *Theobroma* do not share the most recent genome duplication event (Rapp, Udall, and Wendel, unpublished).

3.2 Recent Duplicates

Recently formed allopolyploids, such as *G. hirsutum*, retain duplicated copies of most genes on homoeologous chromosomes. These homoeologous loci typically have sufficiently high sequence identity to obscure the genomic origin of ESTs during molecular and informatic analyses. Because of this technical limitation, the contribution of each homoeolog to a polyploid transcriptome had remained largely unexplored. Recent work indicates that homoeologous transcript levels may not be equal and that unequal homoeologous gene expression is common in allopolyploids (Comai 2000; Kashkush et al. 2002; Kashkush et al. 2002; Osborn et al. 2003; Wu et al. 2003; He et al. 2003; Adams et al. 2003; Adams et al. 2004; Wang et al. 2004; Adams and Wendel 2005; Madlung et al. 2005; Hegarty et al. 2005; Yang et al. 2006; Wang et al. 2006; Lai et al. 2006; Tate et al. 2006; Wang et al. 2006). In both recently synthesized and natural

cotton, differences in homoeologous expression levels of eight genes were detected in which the A-genome copy was expressed more than the D-genome copy or *vice versa* (Adams et al. 2003). The biases were tissue-dependent, with no obvious relationship to inferred gene function. These results suggested an almost immediate initiation of expression subfunctionalization with the onset of polyploid formation, where regulation of gene expression is somehow compartmentalized to different tissue types.

Two subsequent approaches have also documented inequality in the expression level of homoeologous loci. An EST analysis of GH_TMO ovules (Table 1) identified 2,233 genes with only A-genome ESTs in the GH_TMO ovule library and 372 genes with only D-genome suggesting a wide-spread inequality of homoeologous transcript levels in the allotetraploid EST library (Yang et al. 2006). Expression levels of a small sample of these loci were confirmed using SSCP. *In silico* homoeologous transcriptional biases were also found for putative transcription factors and phytohormone signal transduction pathways suggesting subfunctionalization in ovule development and coordinated transcriptional queues among genes from a single genome.

A proof-of-concept, high-throughput study empirically measured transcript abundance providing a limited genomic perspective of this phenomenon using novel, custom Nimblegen microarrays (Udall et al. 2006). Unequal homoeolog expression levels were found for 43% of the 461 surveyed genes in leaf tissue and the expression biases of several genes were also verified using SSCP. In leaf tissue, these genes represented many different functional GO categories and there did not appear to be an over-representation of expressed regulatory transcripts.

Previous work had suggested that a bias in the expression level of homoeologous loci may be due to polyploidization (Adams et al. 2003; Adams et al. 2004). Petal tissue is composed of a simpler cell histology than leaf tissue and a large difference of expression was found between the A-genome and D-genome *AdhA* genes in petal tissue (Adams et al. 2003). Using a custom microarray, the frequency of homoeologous expression biases was found to be very similar between a natural (Acala Maxxa) and re-synthesized (F_1 of $A_2 \times D_5$) tetraploid cotton suggesting that the majority of expression bias between homoeologous loci may be due to inter-specific hybridization (*i.e.* the joining of two distantly related genomes) rather than chromosome doubling (Flagel, Udall, Nettleton, and Wendel, unpublished). In petal tissue, approximately 60% of the detected loci had greater levels of D-genome expression than A-genome expression. Similar to the bias found in ovule tissue (Yang et al. 2006), these results suggest that homoeologous loci partition gene expression by tissue type (*i.e.* subfunctionalization). Several genetic mechanisms have been proposed that may account for different levels of expression, including differential promoter efficiency, histone modifications, DNA methylation, and RNAi (Chen and Ni 2006; Adams 2007), but to date no mechanism has been empirically identified.

4 Identification of Differentially Regulated Genes

Measurement of mRNA transcript levels with Northern blots has been common practice in molecular biology for at least 20 years. Subsequent sequencing-based approaches including suppressive-subtractive hybridization (Straus and Ausubel 1990) have also been successfully used to identify differentially regulated cotton genes, such as those unique to fiber (Talericio and Ray 2001; Haigler et al. 2005). However, large-scale experiments of the *Gossypium* transcriptome are limited on Northern blots by probe number and they are limited in sequence-based approaches by sampling where many targeted sequences may be under-represented or omitted entirely. An alternative approach to empirically measure transcript levels is through DNA microarrays. DNA microarrays have been created for cotton and provide a highly parallelized tool for gene expression investigations in cotton.

4.1 Microarray Development for *Gossypium* Transcriptome Analysis

Microarrays are a powerful method to simultaneously measure relative expression levels for thousands of genes. Traditional microarrays are composed of cDNA inserts, short oligonucleotides, or long oligonucleotides and the advantages and disadvantages of each of these microarray types have been extensively reviewed (Meyers et al. 2004; Alba et al. 2004; Rensink and Buell 2005). The debate of which type is best and/or inconsistencies between platforms has also been addressed elsewhere. Here, I provide a brief historical summary of cotton microarrays and also provide a foundation to those who may wish to use microarrays in cotton gene expression studies. All three types of microarrays have significantly contributed to our understanding of the *Gossypium* transcriptome. The first type of microarray developed for cotton was a cDNA microarray. cDNA microarrays are composed of individual spots of amplified cDNA plasmid inserts (Fig. 2). Similar to other commodity-based research communities, cDNA microarrays were first used to measure gene expression in cotton leaf and root tissues (Dowd et al. 2004). A separate cDNA microarray revealed the impact of ethylene on developing cotton fiber transcript levels (Shi et al. 2006; Gou et al. 2007).

Only a few months after the first cotton cDNA microarray paper, ~13,000 long oligonucleotide probes were designed from an assembly of cotton fiber ESTs (~45,000). This second type of microarray (an oligo-based array) was used to describe differential gene expression between early and late stages of cotton fiber development (Arpat et al. 2004). Independently, a second set of ~13,000 oligonucleotides were synthesized from 180,000 ESTs representing 30 cDNA libraries that sampled an array of tissues and experimental treatments (Udall et al. 2006). Independently, additional oligonucleotides were synthesized

and 13,184 were printed on publicly available microarray slides during 2005 and 2006. Subsequent additions of cotton ESTs to Genbank (> 280,000 ESTs) have been compiled into an assembly of 27,080 contigs and 36,339 singletons (<http://www.ESTinformatics.org>). From the recent influx of EST sequences – particularly from a mature ovule library – an additional complementary 9,000 oligonucleotide probes were synthesized and added to the public long oligonucleotide microarray resulting in 22,787 long oligonucleotides printed on each microarray (Udall et al. 2007).

During the design of the oligonucleotide microarray, there were many possible unigenes in the assemblies to target with oligonucleotide probes. Selection of a subset of possible oligonucleotide targets was constrained by the requirements of having unique probe sequence, a uniform melting temperature (+/- 3 degrees), and scientific community interest regarding subsets of genes such as transcription factors. Detailed information regarding the development of this microarray platform is available (Udall et al. 2007). Researchers using this microarray platform can contribute their results to a growing collection of cotton microarray data at the information portal of Comparative Evolutionary Genomics of Cotton (CEGC, <http://cottonrevolution.info>) or at the Gene Expression Omnibus (GEO) database (GEO#: GPL4808; Barrett et al. 2005). Probe sequence, probe annotation, microarray availability, microarray ordering information, and microarray print-batch quality control information is also available at CEGC.

A third type of cotton DNA microarray, using short oligonucleotide probe sets, is also available through Affymetrix or NimbleGen. Short oligonucleotide probes (25-40 bases) generally target 10 or more independent segments throughout the length of the gene to target individual genes (Fig. 1). The several probes targeting a single gene are called a 'probe set'. On Nov. 10, 2006, Affymetrix announced the release of a commercially available cotton microarray. The GeneChip[®] Cotton Genome Array is a single GeneChip brand array comprised of 23,977 probe sets representing 21,854 cotton transcripts. Gene sequences used in the design of the array were selected from GenBank, dbEST, and Ref-Seq. Individual probe sets for these genes were created from the *G. hirsutum* UniGene database (Build 2, August 2006) and the *G. raimondii* UniGene database (Build 2, September 2005).

Increasingly powerful innovations in microarray technology have been applied to non-model plant genome analysis. Custom microarrays can be designed for gene expression experiments and are based on either short oligonucleotide probe sets (Nimblegen) or single long oligonucleotide probes (probes up to 60 bp long, Agilent). Both companies offer software for in-house array design and provide empirically derived, default options. In these cases, individual researchers bear the responsibility of correct array design. Taliencio and Boykin (Taliencio and Boykin 2007) used custom Agilent arrays to identify genes involved in cotton fiber initiation in a traditional transcript profiling experiment. In addition to traditional transcript profiling, Nimblegen

microarrays have been used to diagnose expression level between the A_T and D_T genomes (Udall et al. 2006).

In addition to DNA microarrays, sequence-based approaches (differential display and suppressive-subtractive hybridization) have been used to characterize the *Gossypium* transcriptome. Most research to date using either DNA microarrays or sequenced-based methodologies has focused on the fiber transcriptome.

4.2 Fiber Transcriptome

While other agricultural and biological aspects of cotton may also be regulated by transcription, a description of the *Gossypium* transcriptome must include a description of the fiber transcriptome because of fiber's economic and research importance. The cotton fiber transcriptome is an excellent model for cellular development and elongation (reviewed by Basra and Malik 1984) and there is a great potential of modifying or bioengineering fiber quantity and quality for cotton production. Proper gene expression is essential for fiber development and different sets of genes are expressed during the four stages of fiber development (Fig. 6), initiation, primary wall synthesis (elongation), secondary wall

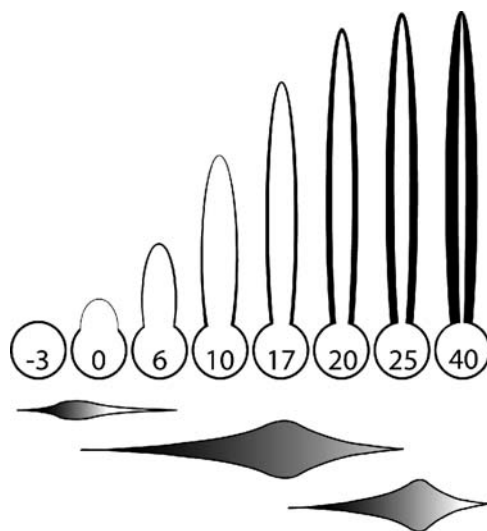


Fig. 6 Developmental stages of single cotton fiber. Numbers indicate days post anthesis. Tear drop shapes beneath the iconographic fiber represent distinct transcriptional stages for fiber initiation, fiber elongation and primary cell wall synthesis, and secondary cell wall synthesis. The gradient in each shape represents a shift in gene expression within a developmental stage (*i.e.* from transcription factors to cellular components) and the shapes are not drawn to scale other than their approximate duration (adapted from (Haigler et al. 2005))

synthesis, and maturation. To a degree, each stage can be considered as having its own distinct transcriptome gradient where transcription factors initiate expression of stage-specific genes, the expression of enzymatic and cellular genes gradually increase and their functional role defines the developmental stage, and finally, any negative transcriptional regulators for each stage are expressed. However, when the transcriptome is assessed at a particular time point, all expressed genes are measured and the various developmental stages have considerable overlap. Several recent studies have identified genes expressed in cotton fiber and quantified their expression. Here, I discuss the relevance of these findings to the *Gossypium* transcriptome.

4.3 Fiber Initiation

Genetic control of fiber initiation is complex and it has been considered key to biotechnological manipulation of cotton fiber (Applequist et al. 2001; Kim and Triplett 2001; Li et al. 2002; Ji et al. 2003; Arpat et al. 2004; Lee et al. 2006). Comparisons between wild-type and naked seed mutants have been particularly useful to identify genes putatively involved in fiber initiation (Yang et al. 2006; Lee et al. 2006; Wu et al. 2006). The transcription signal for fiber initiation is probably very brief for any one epidermal cell when compared to the duration of other fiber development stages. Once that initial signal has been perceived, it triggers a transcriptional cascade of additional downstream transcription factor proteins that begin to modify cell growth. Thus, regulatory genetic components of fiber initiation and those of early cell elongation cannot be entirely separated. Regulatory genes involved in early fiber development have been individually identified as transcription factors and when considered with the results from microarray experiments, they provide a global perspective on fiber initiation.

Many similarities exist between fiber initiation of cotton seed trichomes and *Arabidopsis* leaf trichomes (see Comparative Genomics of Cotton and *Arabidopsis* (Rong), this volume). One of the first types of transcription factors found to control trichome development of *Arabidopsis* were genes from the MYB gene family (Oppenheimer et al. 1991; Hulskamp et al. 1994). Researchers systematically pursued the characterization of MYB transcription factors in cotton because of their role in *Arabidopsis* trichome development (see Wang et al. 2004 for review). Indeed, 6 and 55 cotton genes with MYB domains were found to be expressed during fiber initiation and elongation, respectively (Loguericio et al. 1999; Suo et al. 2003). *GaMYB2* was implicated as an important regulator of fiber initiation because of its transcriptional control of genes preferentially expressed in cotton fiber initials (*LTP3*, *RD11*, and possibly membrane bound *O*-acyl transferase, Liu et al. 2000; Li et al. 2002; Wang et al. 2004; Hsu et al. 2005), its apparent complementation of the *g11* mutant (glabrous) phenotype, and its temporal (early) and spatial expression pattern (Wang et al. 2004). Subsequently, Suo et al. (2003) demonstrated that a different MYB gene, *GhMYB109*, was specifically expressed in cotton fiber initials and in

elongating fibers using *in situ* hybridization, though its specific function has yet to be determined.

Using a microarray approach, Wu et al. (2006) also identified MYB genes that may have a role in fiber initiation. After testing allele complementation between six fiberless or reduced fiber mutants, they compared expression level differences of each genotype's ovule and identified between 60 and 243 genes that were up- or down-regulated in whole 0 days-post-anthesis (dpa) ovules. With this list of genes, the question remained as to which of these differentially regulated genes were expressed in fiber initials as opposed to being expressed in the remaining ovule. The *Gossypium* transcriptome of the ovule outer integument (enriched with fiber initials) was isolated using Laser Capture Microdissection (LCM) and it was compared to the transcriptome of the remaining ovule through microarray hybridization. Among the genes that were differentially regulated in fiberless mutants, only a small number (13) had elevated expression levels in wild-type ovules when compared to the fiberless mutants containing four putative transcription factors, *GHMYB25* (MIXT-type MYB gene), *GhMyb25-Like* (also MIXTA-type MYB gene), *GhHDI1* (homeo-domain gene), and *GhCycD3* (cell-cycle gene, likely involved in endoreduplication during fiber growth; Wu et al. 2006).

Additional types of transcription factor proteins have been found to be involved in fiber initiation (Wu et al. 2006; Yang et al. 2006; Taliercio and Boykin 2007), similar to reports in *Arabidopsis* (Serna and Martin 2006). Yang et al. (Yang et al. 2006) sequenced approximately 32,000 ESTs from a library of -3 to +3 dpa ovules (GH_TMO). In an assembly of all the ESTs in GenBank including four other 'fiber' libraries, they identified 648 transcripts that were over-represented in these libraries through a clustering approach. Several clusters had more abundant ESTs from the GH_TMO library than from the other four libraries and contained several putative transcription factors including MYB, MIXTA-type MYB, MADS, C2H2, C3H, bHLH, and WRKY domains. The transcript abundance in early stages of fiber development was confirmed for 12 selected ESTs using quantitative RT-PCR. Comparing the wild-type transcription abundance to the transcription level of the naked seed mutant, they found a significant reduction in transcript abundance for all 12 genes suggesting that the *in silico* analysis had identified many putative transcription factors involved in fiber initiation and fiber elongation (*i.e.* primary cell wall synthesis). Their list of transcription factors provides several testable hypotheses regarding the genetic control of cotton fiber initiation.

Finally, Taliercio and Boykin (Taliercio and Boykin 2007) recently confirmed the expression of many previously described genes using a custom Agilent microarray in combination with an innovative RNA extraction methodology. In addition to independently confirming other reports, two additional genes were identified in fiber initiation. Both of these genes were *CAPRICE/TRIPTYCHON* (CPC) gene family members and components of a Ca^{+2} mediated signally pathway (calmodulin genes). They were identified as having lower and higher gene expression levels when fiber initials were compared to

10 dpa fibers, respectively. In *Arabidopsis*, *CAPRICE/TRIPTYCHON* acts as a transcription factor that inhibits trichome differentiation on leaves.

In all three studies of the fiber initiation transcriptome, the transcriptional machinery of *Gossypium* seed trichomes shares conserved functional roles with the trichomes of other plants (Yang et al. 2006; Wu et al. 2006; Taliercio and Boykin 2007). To a large degree, the fiber initiation transcriptome appears to have strong similarities to that of *Arabidopsis* where a transcriptional complex containing protein transcription factors such as *GaMYB2* (an *Arabidopsis* GL1-like MYB gene), *HOX3* (a *Arabidopsis* GL2-like HOX gene), *GhMYB109* (a bHLH domain protein; Suo et al. 2003), and *TTG1* (two WD-repeat proteins; Humphries et al. 2005; Taliercio and Boykin 2007) regulate a suite of genes essential for fiber development. The *Gossypium* transcriptional machinery also has similarities to other plants including *Antirrhinum majus* (Noda et al. 1994; Perez-Rodriguez et al. 2005) and *Solanum dulcamara*, (Glover et al. 2004) in which MIXTA-type MYBs have been found to impact trichome development, although these have no known role in trichome development of *Arabidopsis* (Glover et al. 1998; Payne et al. 1999). In fact, seed trichomes have arisen multiple times in different plant families and many species may have functional homologs of seed trichome initiation. Thus, the evolutionary relationships and hypothesized interactions between identified transcription factors remains to be empirically determined (Serna and Martin 2006). Based on the current view of the fiber transcriptome and despite many ‘unknown’ genes whose function remains to be characterized, it appears that a large degree of homologous gene function has been conserved across angiosperm taxa constituting a common molecular program of trichome initiation and subsequent transcriptional cascade that includes cotton.

4.4 Cell Elongation and Primary Cell Wall Formation

Transcription factors that control the initiation of fiber development induce the expression of many genes needed for cell growth, elongation, and primary cell wall synthesis (Fig. 6). Studies of genes involved in fiber elongation suggest that water-mediated cell turgor is the primary driving force behind cell elongation (Ruan and Chourey 1998; Ruan et al. 2001; Ruan 2007). Turgor in the developing fiber vacuole is largely mediated by osmosis. A concentration gradient across the plasma membrane allows water to flow in through gated plasmodesmata (Ruan 2007) and an osmolyte concentration gradient is actively created by vacuolar and plasma membrane H^+ -ATPases (Hasenfratz et al. 1995; Smart et al. 1998), sucrose synthase (Ruan and Chourey 1998), and sucrose and K^+ transporters (Ruan et al. 2001). Results from functional genomic studies largely support these findings and have identified additional genetic factors involved in fiber development.

Unlike fiber initiation, few mutants of cell elongation and primary cell wall formation exist and very few genes have been empirically characterized. One example of a gene effecting cell elongation has been recently described (Shi et al. 2006). *1-Aminocyclopropane-1-Carboxylic acid Oxidase* (ACO) genes 1-3 were found to impact fiber elongation. ACO genes are responsible for ethylene production and were found to be highly expressed during cell elongation. Availability of ethylene appears to promote the expression of other genes essential for fiber elongation such as sucrose synthase, tubulin, and expansin. While the precise function of these induced genes has not yet been fully characterized, many sequences of genes expressed during cell elongation and primary cell wall formation were identified through the first large scale EST sequencing project of cotton (Arpat et al. 2004).

Many gene sequences involved in fiber development were derived through a large-scale EST sequencing project of 7-10 dpa cDNA clones derived from *G. arboreum* fibers (Arpat et al. 2004). From a collection of ~46,000 ESTs, approximately 14,000 EST consensus sequences were identified. An *in silico* analysis of these ESTs determined that the majority of these transcripts were composed of three major functional groups: cell wall structure and biogenesis, cytoskeleton, and energy/carbohydrate metabolism. Along with many unannotated genes, some of the most highly expressed genes in this library included two alpha-tubulin, an aquaporin, and E6 (John and Crow 1992). Using the PAVE browser with updated BLAST comparisons, several unique transcripts were expressed in this particular library including several lipid transfer proteins, aquaporins, arabinogalactin proteins, and proline-rich proteins.

Fiber elongation and primary cell wall formation have been studied independently using different DNA microarray platforms (Ji et al. 2003; Arpat et al. 2004; Lee et al. 2006; Shi et al. 2006; Gou et al. 2007; Hovav, Udall, Hovav, Rapp, Flagel, and Wendel, unpublished). These investigations have illustrated the importance of plant hormones to fiber development, such as ethylene, auxin, and gibberellins that may regulate large suites of genes (Wang et al. 2004). Lists of individual fiber elongation genes were published that include arabinogalactin proteins (AGPs), expansins, tubulins, lipid transfer proteins, proline-rich proteins and chromatin associated genes. To compare gene lists between the six studies, each gene was queried to a common reference, the ESTinformatics EST assembly. The ESTinformatics name of the top BLASTN hit was used as a reference name for each gene. The intersection of these lists identified 98 that were in three or more studies and whose GO terms had significant over-representation of cell elongation genes such as epidermal cell differentiation, cuticle biosynthetic process, and very-long-chain fatty acid metabolic process. The union of these lists contained a total of 4,071 different unigenes. These genes were annotated by GO terms that were significantly over- (147) and under-represented (54) when compared to the entire assembly. In addition to previously mentioned GO categories, the under-represented group included GO factors involved in DNA transcription. The over-represented group included several terms involving water transporter activity and

chromatin assembly or disassembly. Perhaps the regulatory signals of genes annotated with the chromatin associated GO categories somewhat overlaps with cell growth and elongation regulatory signals resulting in additional cellular DNA content. Indeed, endoreduplication has been found in elongating cotton fiber cells (Van't Hof 1999).

4.5 Secondary Cell Wall Formation and Maturation

Secondary cell wall formation is essential to fiber strength, a key quality component of cotton fiber. Once the fiber cell has sufficiently elongated (Fig. 6), cellulose is deposited in massive amounts, thickening the cell wall. Recent reviews have described the physiology and biochemical components of secondary wall formation (Haigler et al. 2005; Ruan 2007; Xu et al. 2007).

Similar to the *in silico* analysis of 7-10 day ESTs, ESTs of a suppressive subtractive hybridization library (GH_SCW, Table 1) were analyzed *in silico* (Haigler et al. 2005) identifying diverse genes related to cellulose biogenesis such as cellulose synthase, chitinase-like gene (Zhang et al. 2004), and sucrose synthase. Comparisons of GH_SCW to the other libraries in PAVE identify versions of beta-tubulin, lipid transfer proteins, cellulose synthases, fasciclin, and hydrolases that are not found in other PAVE cDNA libraries.

Three functional genomics studies also investigated the gene expression of secondary wall formation using DNA microarrays (Arpat et al. 2004; Gou et al. 2007; Hovav, Udall, Hovav, Rapp, Flagel, and Wendel, unpublished) though gene sequences of only two were available in Genbank. As before, gene lists of two studies were compared using the ESTinformatics assembly as a reference. Of 1,165 total genes identified in both studies only 16 were identified in both (too few to test for enrichment of GO terms). This number may be low because 1) there may have been different growing conditions and fiber maturity between the two studies, 2) only the top BLASTN hit was used as annotation while protein function was likely similar for most of the top hits. Nevertheless, the types of genes identified in both studies include expansin, endoribonuclease, cellulose synthase, calmodulin, chitinase, hypothetical proteins, and others. Additional studies of the fiber transcriptome will further develop a consensus picture of secondary wall formation and indeed, all stages of development.

5 Conclusion

The *Gossypium* transcriptome is a new frontier and portions of its landscape are being quickly explored using genomic tools and models extrapolated from closely related *A. thaliana*. The portion that includes the fiber transcriptome has been most extensively explored. Extrapolation of trichome initiation

models has been successful in identifying cotton genes and investigating their function, but many exciting discoveries remain to be found. For example, most of this chapter has focused on protein coding genes; however, significant amounts of the genome may be transcribed but not translated into proteins such as the recent findings described by the ENCODE Project Consortium within 1% of the human genome sequence (The ENCODE Project Consortium 2007). They showed that the majority of the human genome is pervasively transcribed, transcriptional regulatory regions are as likely downstream as upstream, and there is a close relationship between transcription and chromatin assembly and histone modification. Many such insights into the *Gossypium* transcriptome will be illuminated with a sequenced *Gossypium* genome. As a beginning, some non-protein encoding *Gossypium* ESTs have been found to encode microRNAs. Thirty-seven microRNAs of *Gossypium* were computationally identified in the public domain ESTs and appear to target transcription factors and other genes (Qiu et al. 2007).

Gene expression in cotton fiber was the first landscape of the *Gossypium* transcriptome to be surveyed, yet these same genomic tools are also being applied to other aspects of cotton biology and improvement. For example, susceptible response genes were identified as differentially expressed after infection with a root pathogen, *Fusarium infestans* (Dowd et al. 2004). Identified genes were similar to those found in other plants including phytoalexin biosynthesis, other defense related genes, drought responsive genes, other stress response genes (ROS), ethylene signaling genes, and auxin responses, while other cotton gene sequences had no homology to the other organisms. Investigations of abiotic stress tolerance in cotton have begun using candidate genes (as identified in *Arabidopsis* and other plant species; (Light et al. 2005; Huang and Liu 2006; Kosmas et al. 2006; Qu et al. 2006) and one experiment in particular offers a biotechnology improvement to cotton field performance (He et al. 2005). In the future, the *Gossypium* transcriptome will be further characterized as the affect of different treatment conditions are studied empowering further biotechnological advances in cotton fiber production and agriculture.

Acknowledgments A special thanks to Lex Flagel, Ran Hovav, Ryan Rapp, and Jonathan Wendel for proof-reading, assistance with figures, and insightful comments and discussion; and to Foo Cheung for creating the ESTinformatics assembly, proof-reading and helpful discussion.

References

- Adams, K.L. (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* 98, 136–141.
- Adams, K.L., Cronn R., Percifield R., and Wendel J.F. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4649–4654.
- Adams, K.L., Percifield R., and Wendel J.F. (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217–2226.

- Adams, K.L. and Wendel J.F. (2005) Novel patterns of gene expression in polyploid plants. *Trends Genet.* 21, 539–543.
- Alba, R., Fei Z., Payton P., Liu Y., Moore S.L., Debbie P., Cohn J., D'Ascenzo M., Gordon J.S., Rose J.K.C., Martin G., Tanksley S.D., Bouzayen M., Jahn M.M., and Giovannoni J. (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J.* 39, 697–714.
- Amarasinghe, B.H.R.R., Faivre-Nitschke E., Wu Y., Udall J.A., Dennis E.S., Constable G., and Llewellyn D.J. (2006) Genomic approaches to the discovery of promoters for sustained expression in cotton (*Gossypium hirsutum* L.) under field conditions: expression analysis in transgenic cotton and Arabidopsis of a Rubisco small subunit promoter identified using EST sequence analysis and cDNA microarrays. *Plant Biotech.* 23, 437–450.
- Applequist, W.L., Cronn R., and Wendel J.F. (2001) Comparative development of fiber in wild and cultivated cotton. *Evol. Devel.* 3, 3–17.
- Arpat, A., Waugh M., Sullivan J.P., Gonzales M., Frisch D., Main D., Wood T., Leslie A., Wing R., and Wilkins T. (2004) Functional genomics of cell elongation in developing cotton fibers. *Plant Mol. Biol.* 54, 911–929.
- Barrett, T., Suzek T.O., Troup D.B., Wilhite S.E., Ngau W.-C., Ledoux P., Rudnev D., Lash A.E., Fujibuchi W., and Edgar R. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nuc. Acids Res.* 33, D562–566.
- Basra, A.S. and Malik C.P. (1984) Development of the cotton fiber. In: K.W. Jeon (Eds.), *International Review of Cytology*. Elsevier Inc., New York, New York, pp. 65–113.
- Blanc, G. and Wolfe K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678.
- Bowers, J.E., Chapman B.A., Rong J., and Paterson A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438.
- Chen, Z.J. and Ni Z. (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28, 240–252.
- Comai, L. (2000) Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* 43, 387–399.
- Conesa, A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M., and Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.
- Cronn, R.C., Small R.L., Haselkorn T., and Wendel J.F. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* 89, 707–725.
- Cui, L., Wall P.K., Leebens-Mack J.H., Lindsay B.G., Soltis D.E., Doyle J.J., Soltis P.S., Carlson J.E., Arumuganathan K., Barakat A., Albert V.A., Ma H., and dePamphilis C.W. (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749.
- Dowd, C., Wilson I.W., and McFadden H.G. (2004) Gene expression profile changes in cotton root and hypocotyl tissues in response to infection with *Fusarium oxysporum* f. sp. *vasinfectum*. *Mol. Plant-Mic. Inter.* 17, 654–667.
- Endrizzi, J.E., Turcotte E.L., and Kohel R.J. (1985) Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* 23, 271–375.
- Ewing, B. and Green P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194.
- Ewing, B., Hillier L., Wendl M.C., and Green P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.
- Ewing, R.M., Kahla A.B., Poirot O., Lopez F., Audic S., and Claverie J.-M. (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.* 9, 950–959.

- Fulton, T.M., Van der Hoeven R., Eannetta N.T., and Tanksley S.D. (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14, 1457–1467.
- Gaut, B.S. (1998) Molecular clocks and nucleotide substitution rate in higher plants. In: M.K. Hecht (Eds.), *Evolutionary Biology*. Plenum Press, New York, pp. 93–120.
- Glover, B.J., Bunnewell S., and Martin C. (2004) Convergent evolution within the genus *Solanum*: the specialised anther cone develops through alternative pathways. *Gene* 331, 1–7.
- Glover, B.J., Perez-Rodriguez M., and Martin C. (1998) Development of several epidermal cell types can be specified by the same *MYB*-related plant transcription factor. *Devel.* 125, 3497–508.
- Gou, J.-Y., Wang L.-J., Chen S.-P., Hu W.-L., and Chen X.-Y. (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res.* 17, 422–434.
- Haigler, C.H., Zhang D., and Wilkerson C.G. (2005) Biotechnological improvement of cotton fibre maturity. *Physiol. Plant* 124, 285–294.
- Hasenfratz, M.P., Tsou C.L., and Wilkins T.A. (1995) Expression of two related vacuolar H⁺-ATPase 16-kilodalton proteolipid genes is differentially regulated in a tissue-specific manner. *Plant Physiol.* 108, 1395–1404.
- He, C., Yan J., Shen G., Fu L., Holaday A.S., Auld D., Blumwald E., and Zhang H. (2005) Expression of an arabidopsis vacuolar sodium/proton antiporter gene in cotton improves photosynthetic performance under salt conditions and increases fiber yield in the field. *Plant Cell Physiol.* 46, 1848–1854.
- He, P., Friebe B.R., Gill B.S., and Zhou J.-M. (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol. Biol.* 52, 401–414.
- Hegarty, M.J., Jones J.M., Wilson I.D., Barker G.L., Coghill J.A., Sanchez-Baracaldo P., Liu G., Buggs R.J.A., Abbott R.J., Edwards K.J., and Hiscock S.J. (2005) Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Mol. Ecol.* 14, 2493–2510.
- Hsu, C.-Y., Jenkins J.N., Saha S., and Ma D.-P. (2005) Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein. *Plant Sci.* 168, 167–181.
- Huang, B. and Liu J.-Y. (2006) A cotton dehydration responsive element binding protein functions as a transcriptional repressor of DRE-mediated gene expression. *Biochemical and Biophysical Research Communications* 343, 1023–1031.
- Huang, X. and Madan A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Hughes, A. and Friedman R. (2004) Expression patterns of duplicate genes in the developing root in *Arabidopsis thaliana*. *J. Mol. Evol.* 60, 247–256.
- Hulskamp, M., Misra S., and Jurgens G. (1994) Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* 76, 555–66.
- Humphries, J.A., Walker A.R., Timmis J.N., and Orford S.J. (2005) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* *TRANSPARENT TESTA GLABRA1 (TTG1)* gene. *Plant Mol. Biol.* 57, 67–81.
- Ji, S.-J., Lu Y.-C., Feng J.-X., Wei G., Li J., Shi Y.-H., Fu Q., Liu D., Luo J.-C., and Zhu Y.-X. (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nuc. Acids Res.* 31, 2534–2543.
- Jiang, C., Wright R.J., El-Zik K.M., and Paterson A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. U.S.A.* 95, 4419–4424.
- John, M. and Crow L. (1992) Gene expression in cotton (*Gossypium hirsutum* L.) fiber: Cloning of the mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5769–5773.

- Kashkush, K., Feldman M., and Levy A.A. (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160, 1651–1659.
- Kim, H.J. and Triplett B.A. (2001) Cotton fiber growth in planta and *in vitro*. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127, 1361–1366.
- Kohel, R.J., Stelly D.M., and Yu J. (2002) Tests of six cotton (*Gossypium hirsutum* L.) mutants for association with aneuploids. *J. Hered.* 93, 130–132.
- Kosmas, S.A., Argyrokastritis A., Loukas M.G., Eliopoulos E., Tsakas S., and Kaltsikes P.J. (2006) Isolation and characterization of drought-related trehalose 6-phosphate-synthase gene from cultivated cotton (*Gossypium hirsutum* L.). *Planta* 23, 329–339.
- Lacape, J.-M., Nguyen T.-B., Courtois B., Belot J.-L., Giband M., Gourlot J.-P., Gawryziak G., Roques S., and Hau B. (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* x *Gossypium barbadense* backcross generations. *Crop Sci.* 45, 123–140.
- Lai, Z., Gross B.L., Zou Y., Andrews J., and Rieseberg L.H. (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. *Mol. Ecol.* 15, 1213–1227.
- Lee, J., Hassan O., Gao W., Wei N., Kohel R., Chen X.-Y., Payton P., Sze S.-H., Stelly D., and Chen Z. (2006) Developmental and gene expression analyses of a cotton naked seed mutant. *Planta* 223, 418–432.
- Li, C.-H., Zhu Y.-Q., Meng Y.-L., Wang J.-W., Xu K.-X., Zhang T.-Z., and Chen X.-Y. (2002) Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. *Plant Sci.* 163, 1113–1120.
- Li, W.-H., Yang J., and Gu X. (2005) Expression divergence between duplicate genes. *Trends Genet.* 21, 602–607.
- Li, X.-B., Fan X.-P., Wang X.-L., Cai L., and Yang W.-C. (2005) The cotton *ACTIN1* gene is functionally expressed in fibers and participates in fiber elongation. *Plant Cell* 17, 859–875.
- Liang, P. and Pardee A. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–971.
- Light, G.G., Mahan J.R., Roxas V.P., and Allen R.D. (2005) Transgenic cotton (*Gossypium hirsutum* L.) seedlings expressing a tobacco glutathione S-transferase fail to provide improved stress tolerance. *Planta* 222, 346–354.
- Liu, H.-C., Creech R.G., Jenkins J.N., and Ma D.-P. (2000) Cloning and promoter analysis of the cotton lipid transfer protein gene *Ltp3*. *BBA - Mol. Cell Biol. Lipids* 1487, 106–111.
- Loguerio, L.L., Zhang J.-Q., and Wilkins T.A. (1999) Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). *Mol. Genet. Gen.* 261, 660–671.
- Madlung, A., Tyagi A.P., Watson B., Jiang H., Kagechi T., Doerge R.W., Martienssen R., and Comai L. (2005) Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41, 221–230.
- Meyers, B.C., Galbraith D.W., Nelson T., and Agrawal V. (2004) Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiol.* 135, 637–652.
- Noda, K., Glover B.J., Linstead P., and Martin C. (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369, 661–4.
- Oppenheimer, D.G., Herman P.L., Sivakumaran S., Esch J., and Marks M.D. (1991) A *myb* gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67, 483–93.
- Osborn, T.C., Pires J.C., Birchler J.A., Auger D.L., Jeffery Chen Z., Lee H.-S., Comai L., Madlung A., Doerge R.W., Colot V., and Martienssen R.A. (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147.
- Paterson, A.H., Saranga Y., Menz M., Jiang C.-X., and Wright R.J. (2003) QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theor. Appl. Genet.* 106, 384–396.
- Payne, T., Clement J., Arnold D., and Lloyd A. (1999) Heterologous *myb* genes distinct from *GL1* enhance trichome production when overexpressed in *Nicotiana tabacum*. *Devel.* 126, 671–82.

- Perez-Rodriguez, M., Jaffe F.W., Butelli E., Glover B.J., and Martin C. (2005) Development of three different cell types is associated with the activity of a specific *MYB* transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Devel.* 132, 359–70.
- Pertea, G., Huang X., Liang F., Antonescu V., Sultana R., Karamycheva S., Lee Y., White J., Cheung F., Parvizi B., Tsai J., and Quackenbush J. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652.
- Qiu, C.X., Xie F.L., Zhu Y.Y., Guo K., Huang S.Q., Nie L., and Yang Z.M. (2007) Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene* 395, 49–61.
- Qu, Z.-L., Zhong N.-Q., Wang H.-Y., Chen A.-P., Jian G.-L., and Xia G.-X. (2006) Ectopic expression of the cotton non-symbiotic hemoglobin gene GhHbd1 triggers defense responses and increases disease tolerance in Arabidopsis. *Plant Cell Physiol.* 47, 1058–1068.
- Rabinowicz, P.D., Citek R., Budiman M.A., Numberg A., Bedell J.A., Lakey N., O'Shaughnessy A.L., Nacimientto L.U., McCombie W.R., and Martienssen R.A. (2005) Differential methylation of genes and repeats in land plants. *Genome Res.* 15, 1431–1440.
- Rensink, W.A. and Buell C.R. (2005) Microarray expression profiling resources for plant genomics. *Trends Plant Sci.* 10, 603–609.
- Ronning, C.M., Stegalkina S.S., Ascenzi R.A., Bougri O., Hart A.L., Utterbach T.R., Vanaken S.E., Riedmuller S.B., White J.A., Cho J., Pertea G.M., Lee Y., Karamycheva S., Sultana R., Tsai J., Quackenbush J., Griffiths H.M., Restrepo S., Smart C.D., Fry W.E., van der Hoeven R., Tanksley S., Zhang P., Jin H., Yamamoto M.L., Baker B.J., and Buell C.R. (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* 131, 419–429.
- Ruan, Y. (2007) Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fibre. *Func. Plant Biol.* 34, 1–10.
- Ruan, Y.-L. and Chourey P.S. (1998) A fiberless seed mutation in cotton is associated with lack of fiber cell initiation in ovule epidermis and alterations in sucrose synthase expression and carbon partitioning in developing seeds. *Plant Physiol.* 118, 399–406.
- Ruan, Y.-L., Llewellyn D.J., and Furbank R.T. (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K⁺ transporters and expansin. *Plant Cell* 13, 47–60.
- Schlueter, J.A., Dixon P., Granger C., Grant D., Clark L., Doyle J., and Shoemaker R. (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47, 868–876.
- Senchina, D.S., Alvarez I., Cronn R.C., Liu B., Rong J., Noyes R.D., Paterson A.H., Wing R.A., Wilkins T.A., and Wendel J.F. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Bio. Evol.* 20, 633–643.
- Serna, L. and Martin C. (2006) Trichomes: different regulatory networks lead to convergent structures. *Trends Plant Sci.* 11, 274–280.
- Shi, Y.-H., Zhu S.-W., Mao X.-Z., Feng J.-X., Qin Y.-M., Zhang L., Cheng J., Wei L.-P., Wang Z.-Y., and Zhu Y.-X. (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18, 651–664.
- Smart, L.B., Vojdani F., Maeshima M., and Wilkins T.A. (1998) Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. *Plant Physiol.* 116, 1539–1549.
- Straus, D. and Ausubel F. (1990) Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1889–1893.
- Suo, J., Liang X., Pu L., Zhang Y., and Xue Y. (2003) Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochem. Biophysica Acta* 1630, 25–34.

- Taliercio, E. and Ray J.D. (2001) Identification of transcripts translated on free or membrane-bound polyribosomes by differential display. *Plant Mol. Biol. Rep.* 19, 269.
- Taliercio, E.W. and Boykin D. (2007) Analysis of gene expression in cotton fiber initials. *BMC Plant Biology* 7, 22.
- Tate, J.A., Ni Z., Scheen A.-C., Koh J., Gilbert C.A., Lefkowitz D., Chen Z.J., Soltis P.S., and Soltis D.E. (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173, 1599–1611.
- The ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.
- Udall, J.A., Flagel L.E., Cheung F., Woodward A., Hovav R., Rapp R.A., Swanson J.M., Lee J., Gingle A., Nettleton D., Town C.D., Chen Z.J., and Wendel J.F. (2007) Spotted cotton oligonucleotide microarrays for gene expression analysis. *BMC Genomics* 8, 81.
- Udall, J.A., Swanson J.M., Haller K., Rapp R.A., Sparks M.E., Hatfield J., Yu Y., Wu Y., Dowd C., Arpat A.B., Sickler B.A., Wilkins T.A., Guo J.Y., Chen X.Y., Scheffler J., Taliercio E., Turley R., McFadden H., Payton P., Klueva N., Allen R., Zhang D., Haigler C., Wilkerson C., Suo J., Schulze S.R., Pierce M.L., Essenberg M., Kim H., Llewellyn D.J., Dennis E.S., Kudrna D., Wing R., Paterson A.H., Soderlund C., and Wendel J.F. (2006) A global assembly of cotton ESTs. *Genome Res.* 16, 441–450.
- Udall, J.A., Swanson J.M., Nettleton D., Percifield R.J., and Wendel J.F. (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173, 1823–1827.
- Van't Hof, J. (1999) Increased nuclear DNA content in developing cotton fiber cells. *Am. J. Bot.* 86, 776–779.
- Wang, J., Tian L., Lee H.-S., Wei N.E., Jiang H., Watson B., Madlung A., Osborn T.C., Doerge R.W., Comai L., and Chen Z.J. (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507–517.
- Wang, J., Tian L., Madlung A., Lee H.-S., Chen M., Lee J.J., Watson B., Kagochi T., Comai L., and Chen Z.J. (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167, 1961–1973.
- Wang, S., Wang J.-W., Yu N., Li C.-H., Luo B., Gou J.-Y., Wang L.-J., and Chen X.-Y. (2004) Control of plant trichome development by a cotton fiber *MYB* gene. *Plant Cell* 16, 2323–2334.
- Wendel, J.F. (2000) Cotton. In: J. Smartt (Eds.), *Evolution of crop plants*. Longman, London, pp. 358–366.
- Wendel, J.F. and Cronn R.C. (2003) Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186.
- Wilkins, T.A. and Jernstedt J.A. (1999) Molecular genetics of developing cotton fibers. In: AS Basic (Eds.), *Cotton Fibers*. Haworth Press, New York, pp. 231–269.
- Wilkins, T.A. and Arpat A.B. (2005) The cotton fiber transcriptome. *Physiol. Plant* 124, 295–300.
- Wright, R.J., Thaxton P.M., El-Zik K.M., and Paterson A.H. (1998) D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics* 149, 1987–1996.
- Wu, L.M., Ni Z.F., Meng F.R., Lin Z., and Sun Q.X. (2003) Cloning and characterization of leaf cDNAs that are differentially expressed between wheat hybrids and their parents. *Mol. Genet. Gen.* 270, 281–286.
- Wu, Y., Machado A.C., White R.G., Llewellyn D.J., and Dennis E.S. (2006) Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol.* 47, 107–127.
- Wu, Y., Rozenfeld S., Defferrard A., Ruggiero K., Udall J., Kim H., Llewellyn D., and Dennis E. (2005) Cycloheximide treatment of cotton ovules alters the abundance of specific classes of mRNAs and generates novel ESTs for microarray expression profiling. *Mol. Genet. Gen.* 274, 477–493.

- Xu, Y., Li H.-B., and Zhu Y.-X. (2007) Molecular biological and biochemical studies reveal new pathways important for cotton fiber development. *Journal of Integrative Plant Biology* 49, 69–74.
- Yang, S.S., Cheung F., Lee J.J., Ha M., Wei N.E., Sze S.-H., Stelly D.M., Thaxton P., Triplett B., Town C.D., and Chen Z.J. (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J.* 47, 761–775.
- Zhang, D., Hrmova M., Wan C.-H., Wu C., Balzen J., Cai W., Wang J., Densmore L.D., Fincher G.B., Zhang H., and Haigler C.H. (2004) Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. *Plant Mol. Biol.* 54, 353–372.

Genetic Engineering of Cotton

Norma L. Trolinder

Abstract The words biotechnology, genetic engineering, molecular biology, and high throughput genomic analysis have engendered awe, doubtfulness, ambivalence, and hope from scientists and the public alike. The technologies justify the responses, for they are undoubtedly the most powerful biological research tools in existence today. They are not, however, new tools that have suddenly burst upon us. Rather, as with most new tools, the scientific community has been slowly developing them for decades. We must regard them as tools to address and solve real problems not as ends unto themselves. Genetic engineering of cotton has proven to be a very challenging undertaking but despite the challenges, Cotton has led the way for acceptance of genetically engineered crop plants and today the industry stands on a new threshold, with another new set of tools for understanding the cotton genome. It is with that foundation that we move forward today.

1 Introduction

Scientists have sought to understand and manipulate the plant genome since the 1920s when x-rays and colchicine were the leading tools for manipulating chromosomes. It was not until the 1940 – 1950s that scientists determined that DNA was the basis of inheritance and elucidated its structure. Between 1960 and 1970, plant growth regulators were utilized to culture plant cells and tissues in vitro and totipotency was confirmed . Plant cells at relatively early states of development, such as those in quiescent parenchyma, meristems, vascular cambium, and embryonic tissue are undetermined and capable of switching pathways of development in response to environmental stimuli, to reproduce an entire plant or proliferate rapidly to produce undetermined cell masses called calluses, often seen on wounded tissue. The crown gall disease of

N.L. Trolinder (✉)
Texas Tech University & Genes Plus
e-mail: genes@chipshot.net

plants became known to be caused by a soil bacterium, *Agrobacterium tumefaciens*. The bacteria induces rapid proliferation of undifferentiated tissue to form the massive “gall” on the plant. Now we know the structure of DNA, the genetics of DNA, how plant hormones regulate tissue growth and how to regenerate plants from masses of undifferentiated tissue. The various parts of the genetic puzzle were coming together. There was lacking, however, the ability to manipulate the parts .

Studies of the DNA involved in the crown gall disease of plants elucidated the mechanisms whereby the bacterium inserted its own DNA into plant DNA during infection. The seminal work of Marc Van Montagu and Jozef Schell of Belgium, Mary Dell Chilton of the U.S.A., and Robert Schilperoort of The Netherlands led the way in elucidating the nature of the soil bacterium, responsible for the infection. During the 1970-1980 period, enzymes were discovered that could cut and splice DNA at very specific sites in its structure, allowing the first recombinant DNA work. Recombinant DNA techniques were used to splice foreign DNA into and/or remove portions of the crown gall organisms’ own DNA, followed by infection of plant tissue by the bacterium to elucidate the various functions of the organisms genome.

Thus, the potential to move genes between species and phyla and to dissect the structure and function of plant genomes, opened the door to improvement of agronomically important crop plants such as cotton. Central to this improvement is the contribution of plant tissue culture, regeneration of whole plants from plant cells and/or tissues, and transfer of foreign DNA into the plants own DNA, a process now termed plant transformation.

2 Tissue Culture – A Brief History

Totipotency, the ability of a single cell to regenerate a whole organism, was predicted as early as 1878 and somatic embryogenesis in 1902, but it was not until 1927 when the auxin IAA was shown to promote cell growth that progress in plant tissue culture began to move forward. During the 1930s another critical component was added when a defined medium that sustained growth of isolated plant roots and undifferentiated callus tissue was developed. The discovery that kinetin promoted cell division in 1957 instigated studies leading to understanding the interaction of cytokinins and auxins on root and shoot differentiation from callus (ie. organogenesis) and micro-propagation (multiple shoot formation from apical and lateral buds). Micro-propagation became an important tool in the plant nursery industry and today is an important tool for plant transformation. This was quickly followed by the discovery of somatic embryogenesis in carrot callus. Somatic embryogenesis and organogenesis became the primary tissue culture tools for transformation of agronomic crops.

Undifferentiated callus tissue can be transferred to liquid culture medium and agitated to form plant cell suspensions and maintained over an extensive

period of time. Cells derived from such suspensions can be regenerated by organogenesis or somatic embryogenesis. Cell suspension cultures are amenable to selection against various biotic and abiotic stresses, generating mutant populations from which whole plants can be recovered. Indeed, much like bacterial cell cultures, continuous culture of plant cell suspensions in bioreactors can be utilized to produce biochemical products in response to stimuli or during the natural course of the cell cycle.

Plant protoplast culture was made possible after their release from root-tip cells by a fungal cellulase was observed in early 1960. Protoplasts may be isolated from intact plant tissues such as leaf, root tips, immature pollen, or tissue cultured cells, preferably fine cell suspensions. The value of protoplast culture lies in enabling studies of the plasma membrane, fusion of the plasma membranes of different cell types, organelles, and/or species (somatic hybridization), and the uptake of large fragments of DNA via liposomes. Liposomes are artificial vesicles prepared from phosphatidyl choline and stearyl-amine that can protect nucleic acids encapsulated in the vesicals from cellular nucleases after fusion with the plasma membrane of protoplasts.

It is evident that the 1950s and 1960s saw a plethora of advances in plant tissue culture, including the first callus formation from gymnosperm and angiosperm pollen grains. In 1964, cultured angiosperm pollen was induced to form embryoids that developed into whole plants, thus providing a valuable method for producing haploid plants. Regeneration from microspores and the female gametophyte has been accomplished in several gymnosperms as well as angiosperms. Although most plant tissues require growth regulator treatments to induce organogenesis or embryogenesis, some tissues, such as ovules and pollen, may not require any added growth regulators. Other tissue explants such as epidermal strips, flower peduncles, leaf bases, petioles, and mid-veins may be in a pre-existing physiological state that allows them to produce flowers, adventive embryos or adventitious shoots directly without going through a callus stage.

2.1 Tissue Culture of Cotton

The first application of tissue culture techniques for cotton improvement was an attempt to induce nodal segments and shoot apices from field grown plants to form roots *in vitro* for the purpose of germplasm preservation (Chappell and Mauney, 1967). It was not until several decades later that *in vitro* rooting became an issue once again. This time emphasis was on the germination and rooting of somatic embryos of cotton or the plants generated thereby (Gould et al, 1991). Subsequently it became important in the recovery of shoots formed from putatively transformed meristem tissue (Luo & Gould, 1999).

Cotton callus formation was first reported by Beasley (1971). Beasley's attention was focused on the use of tissue culture techniques to study cotton fiber development. During his studies he observed callus formation on

immature cotton ovules. Callus formation was not his intent and it remained to others to continue with studies related specifically to cotton callus. However, Beasley successfully demonstrated that the cotton ovule could be induced to initiate and sustain fiber development *in vitro*. Subsequently, Beasley and Ting (1973) demonstrated that this method also supported the development of embryos in 2 DPA ovules. Stewart and Hsu (1977) used ovule culture to rescue interspecific hybrid embryos. Today, Beasley's method of *in vitro* culture of cotton ovules continues to be an important tool in the study of fiber growth and development from biochemical, morphological and gene action perspectives (Kim and Triplett, 2001).

Callus formation on immature ovules was utilized by Trolinder and Goodin (1987) to demonstrate that suspension cultures derived from immature cotton ovule callus could be induced to elongate and form a limited amount of secondary cell wall. These cells were maintained in a proliferative state in air lift bioreactors with continuous nutrient supplementation for very prolonged periods and still maintained the capacity to elongate when induced by gibberellic acid. Elongated cells form a limited amount of secondary cell wall. It is of interest that the degree of fiber elongation and secondary cell wall deposition diminishes with each step of removal from the *in planta* state. Fibers produced when the Beasley method is utilized, have a reduced length and secondary cell wall deposition as compared to those produced *in planta*. Induced cell suspensions derived from cotton ovule callus have reduced elongation and secondary cell wall deposition compared to those produced by the Beasley method. This factor could have further implications in the study of genes associated with fiber cell elongation and cell wall deposition.

Subsequent to the work of Beasley, several investigators described conditions for the establishment and growth of cotton callus and cell suspension cultures (Price and Smith 1977). Barrow (1986) established a culture medium that sustained viable cotton microspores but was unsuccessful in obtaining plants. To date haploid cotton plants have not been regenerated from pollen. The first report of somatic embryogenesis in cotton was made by Price and Smith (1979). The species was a wild cotton, *Gossypium klotzchianum*. Although somatic embryos were obtained, the investigators were unable to germinate the embryos and the technique did not seem to be applicable to cultivated cottons. However, much was learned from these seminal investigations that ultimately produced increasing success with cultivated cottons (Davidonis and Hamilton, 1983; Finer & Smith, 1984; Rangan et al, 1984; Robacker et al., 1984; Trolinder and Goodin, 1985, 1987, 1988a, 1988b; Shoemaker et al., 1986; 1997; Finer, 1988; Gawel & Robacker, 1990; Sun et al, 2005).

Further investigations uncovered possible limitations to the successful use of somatic embryogenesis in germplasm improvement. The cultural procedures described by these investigators could not be uniformly applied to all cultivated cottons, indicating a recalcitrance of certain genotypes to regeneration by somatic embryogenesis (Trolinder and Xhixian, 1989). Highly embryogenic Coker lines, Coker 312 and Coker 5110 were identified. Further screening of

Coker lines identified Coker 310, 201, 5110, 312, and 315 as embryogenic as well. These lines were the products of a cross between Coker 100 W and Delta Pine 15 (Lloyd Langford, SeedCo, personal communication). Reconstitution of this cross identified 100 W as the progenitor of the embryogenic trait in these lines (Trolinder, unpublished data). Other lines derived from this background are potential candidates for regeneration via somatic embryogenesis. Extensive screening by this author and others has identified many responsive cultivars, including, but not limited to GC510, Pee Dee germplasm, Mar(Multi-Adversity Resistance) germplasm, Sure Grow 501, DP90, Texas Racestock T25, Chinese variety Lu 1, and Acala lines. Although a number of cultivars have indeed been proven regenerable by somatic embryogenesis only a limited number yield the frequencies desired for efficient transformation. However, individual plants within a recalcitrant cultivar may be identified as having efficient embryogenic potential. These plants may be propagated as a germplasm source for *in vitro* genetic manipulations.

The literature is replete with variations of culture conditions to expand the scope of regenerable cotton varieties (Rangan & Zavala, 1984; Shoemaker et al, 1986; Firoozabady, 1993; Zhang et al, 2001; Mishra et al 2003; Jim et al. 2006, Wang et al, 2006; Sun, 2006). Research continues to be focused on genotype independent regeneration of cotton including the study of genes associated with somatic embryo-genesis (Gawel and Robacker, 1990; Zeng et al 2006). Trolinder et al. (1999) were able to increase the number of recalcitrant cultivars forming somatic embryos by taking advantage of the natural gradient of hormones in the germinating embryo. The transition region of hypocotyl accumulates both cytokinin and auxin as a result of *in vivo* gradients that form during germination. Thus when this tissue is utilized as the explant, no additional hormone, synthetic or natural, is required and somatic embryos rapidly appear directly on the initial calli.

While at first look, genotypic dependence of cotton somatic embryogenesis appeared to be a stumbling block to large scale commercial use, it is certainly not an insurmountable one, nor even of primary concern in the commercial realm. In today's commercial environment where regeneration of transformed somatic embryos leads to a commercial product, only one or two transformation events producing a specific product will go through the regulatory hoops necessary for commercialization. The cost of such regulatory procedures is enormous. Similarly, the process of transformation and regeneration of multiple cultivars would be prohibitively expensive. Today, the plants carrying elite transformation events are back crossed into relevant cultivars to produce a commercially viable product. Conversely, to utilize transformation as a tool for cotton genomics, it would be desirable to be able to transfer genes readily in any background.

Efficiency of conversion of somatic embryos to whole plants is of concern for public and private researchers alike. Cell cultures are mostly asynchronous with many different stages of embryo development present in any particular culture. Thus, of thousands of somatic embryos present in a culture, only a few may be

at similar stages of development at any one time. Since nutrient availability is continuously declining in a closed culture, most of the smaller embryos may never be recovered, as transfer to fresh media stimulates proliferation of embryos rather than maturation. This may be somewhat overcome by optimizing density during development. In general, transferring cells to fresh medium, solid or liquid suspension, at low density promotes development and high density promotes proliferation (Trolinder and Goodin, 1988b). Another approach is to better reproduce conditions present during normal *in planta* embryo development through culture media and environmental conditions.

Attempts to synchronize cultures in order to recover a greater percentage of somatic embryos, include sieving of embryogenic cell suspensions prior to plating on embryo development medium (Trolinder and Goodin, 1987; Wilkins et al, 2004, Cao et al, 2008), and alteration in pH (Shang & Trolinder, 1991). Low temperature treatment of cells is a potential method for synchronization. Cells exposed to sub-optimal temperature complete the current mitotic activity but will not enter into a new cycle of mitosis until the stress is removed. For optimum synchronization, single cell cultures would be most amenable. The combination of sieving to obtain a very fine suspension consisting primarily of single highly cytoplasmic cells and low temperature treatment might be promising in this case. The downside to this approach is a more extended time in culture to achieve the required fine fraction. The extended time in culture can in itself present other problems, primarily an increase in somaclonal variation (Jain et al., 1998).

Somaclonal variation, the inherent variability produced during cell culture, is a limitation to efficient use of somatic embryogenesis and organogenesis. Undesirable variability in ploidy level, chromosomal bridges at anaphase, multi-polar spindles, lagging chromosomes, fragments and unequal separation during anaphase is common in tissue-cultured cells (Bayless, 1973). This has proven true for cotton (Li et al. 1989; Bajaj and Gill, 1985; Stelly, et al. 1989). Polyploidy and/or aneuploidy may have pre-existed in the tissue explants or been induced by the culture process itself. Synthetic auxins such as NAA and 2,4-D are known to induce spindle failure and other mitotic abnormalities. Extended time in culture increases the extent of variability. Initially, there was interest in producing variability in cotton cell culture that could be utilized in germplasm development. Cells were subjected to rigorous selection against biotic and abiotic stresses such as pathogens (Rangan et. al., 1998, 1999) and temperature (Trolinder and Xhixian, 1991). While such variations may be utilized in a breeding program, they are not applicable when a true to type plant is desired. Over a period of years, it has been observed that the speed and efficiency with which cotton somatic embryos can be produced has a profound effect on the observed phenotypic variability of regenerated plants. The shorter the time in culture and/or exposure to growth regulators, the less variability is observed (N. Trolinder, unpublished data)

The importance of obtaining true to type plants from somatic embryos of cultivated cottons became more apparent with the advent of the ability to

transfer foreign genes into the cotton genome, i.e. transformation. With the goal of obtaining a specific heritable genotype, genotype independent regeneration, and reduced variability in regenerated plants, focus was placed on alternate methods of regeneration that reduce or eliminate somaclonal variability. These methods include recovery of plants by organogenesis, the shoot apex, and axillary buds.

Organogenesis, the *de novo* organization of cells into meristematic buds, gives rise to organized tissue, stems, leaves, roots, and flowers, individually or as entire shoots. Somatic embryos are known to develop from single cells, whereas plants produced by organogenesis can potentially be derived from multiple cells and thus be chimeric. In truth, however, few plants derived via organogenesis from other species, such as tobacco, have been shown to be chimeric. Reichert et al (2002) reported regeneration of cotton via organogenesis. To determine true organogenesis, one must observe carefully as often the explant may actually contain pre-existing nodal regions that produce axillary buds (micro-propagation) rather than *de-novo* regeneration. Micro-propagation is the breaking and multiplication of axillary buds and has the potential to provide hundreds of identical plants from one nodal region.

Regeneration of the shoot apex eliminates undesirable somaclonal variation but normally provides only a one to one return, i.e. one apex = one plant, whereas micro-propagation or somatic embryogenesis provides a number of explants per seedling and hundreds of plants per explant. Bajaj and Gill (1986) followed by Gould et al (1991), were first to report successful isolation and culture of the intact cotton shoot apex. Recovery of shoots from dissected halves or quarters of the shoot apex plus cotyledonary nodes was reported, eliminating the one to one recovery rate (Trolinder et al, 2006). A number of other reports of recovery of whole cotton plants from the shoot apex have surfaced with varying degrees of success. Agrawal et. al. (1997, Tuli et. al. (2001), Zhong et. Al. (2005), and Huzera et al (2007) reported successful micro-progation of the shoot apex and axillary buds. Attempts to root shoots recovered from the shoot apex, nodal segments, or organogenesis were problematic for some investigators (Luo & Gould, 1999; Ouma et. al., 2004) whereas others (Trolinder et. Al., 2006) found rooting was accomplished without complex manipulation or medium.

3 Transformation of Cotton – the Techniques

Agrobacterium mediated delivery of transgenes to cotton hypocotyl tissue, followed by callus formation, somatic embryo development and plant regeneration has been the primary method of choice for cotton transformation. The plant pathogen *Agrobacterium tumefaciens* harbors a plasmid which has the natural ability to transfer a segment of its' own DNA (T-DNA) into the plant genome. Segments of the T-DNA that cause the plant tissue to produce a

compound needed by the bacteria, can be removed and replaced with foreign DNA of interest. Transgenic plants regenerated from infected tissues contain the foreign DNA of interest. The gene of interest segregates in a Mendelian fashion in progeny of the transgenic plant (Hood et al, 1999).

The ability to transform plant tissues and regenerate entire plants from the transformed tissue was initially restricted to those species and genotypes that could be easily regenerated in tissue culture following the model of Horsch et al (1985). For Horsch's organogenic model, tobacco, regeneration was a "slam dunk". Unfortunately, cotton did not fall into that category and regeneration was for some time a stumbling block to its successful transformation. Therefore, transformation of cotton lagged behind that of model dicot species such as tobacco and carrot. Umbeck et. Al. (1987, 1991, 1992) published the first report of cotton transformation coincident with the publications of Trolinder and Goodin (1985, 1987) detailing regeneration of somatic embryos of cotton. This collaboration resulted in a seminal cotton transformation patent for Agracetus and Trolinder (U.S. patent nos. 5,004,863 and 5,159,135). Following these reports, there was rapid dissemination of the information needed to successfully transform and regenerate cotton. Other investigators added substantially to the knowledge base for cotton transformation and regeneration (Cousins et al., 1991; Bayley et al., 1992; Firoozabady, 1993; Strickland, 1998; Kumar et. al., 1998; Hemphill et. al., 1998; Rangan et al., 2003; Misra et al., 2003; Raeynaerts & Sonville, 2002).

A number of other methods of DNA delivery have been investigated with varying degrees of success. Klein et al. (1987) introduced high-velocity micro-projectiles as the carrier of DNA into cells. DNA fragmentation during bombardment can result in incorporation of fragments as well as full length genes and/or multiple copies of the gene of interest (Klein et. al., 1987). Similar patterns are observed with *Agrobacterium* transformation. Chlan et.al. (1995), and McCabe and Martinelli (1993) were first to successfully utilize gold particles coated with naked DNA as micro-projectiles to bombard meristematic tissue of germinating cotton embryos. For cotton, gold particles coated with DNA must penetrate beyond the L1 & L2 tissue layers to the L3 layer of meristems to obtain germline transformants. While this method has the advantage of eliminating somaclonal variation and genotype dependence of somatic embryogenesis, the frequency of germ-line transformants is quite low, 0.06 to 0.14 %. Initial transformants are chimeric, necessitating extensive screening via a reporter gene followed by selective pruning of identified nodes to obtain germline transformants. Micro-projectile bombardment is also useful for obtaining plants free of antibiotic resistance genes and transfer of multiple genes. Visible markers may be used to identify transformed areas containing inserts of the differently tagged genes. Finer et al (1990) successfully used particle gun bombardment of embryogenic cotton cell suspension culture to obtain transgenic cotton plants

Transformation by *Agrobacterium* infection of the cotton shoot apex were first reported by Ulian et. al., (1988) but transformation frequency was low. To improve frequency of germline transformation of the shoot apex, methods were

developed to provide greater access to germline tissues. Trolinder et al (1999) exposed a greater number of germ-line cells to the transforming agent by splitting the shoot apex into halves and/or quarters. Plants were successfully recovered at high frequency. To further increase the number of receptive germ-line cells exposed to the transforming agent, apices were exposed to low temperature prior to and during splitting. Similarly, Jiang (2004) attained transgenic cotton plants by wounding shoot apices to allow greater access of *Agrobacterium* to the L3 layer. Hussain et. al. (2007) subjected meristematic tissues to sonication to form cavitations through which *Agrobacterium* may reach germline layers. To date, all methods of *Agrobacterium* mediated transformation of meristems suffer from the same drawbacks as particle gun transformation in terms of frequency, chimeric nature, and selection of transformants.

In vivo transformation of cotton has been attempted by several laboratories. A method of particular interest in China has long been *in planta* injection of developing cotton ovules with DNA. Zhou et. al. (1983) claimed transformation of one day post anthesis ovules by injecting naked DNA into the axil placenta. They suggested that the DNA transformed the embryos by entering the ovule through the pollen tube pathway. They found variation in phenotypes when DNA of *G. barbadense* was introduced into glandless *G. hirsutum* embryos, Sea Island DNA into *G. arboreum* and *G. hirsutum*, *G. thurberi* into Upland Cotton, *Abutilon avicennae* into Sea Island and a *Fusarium*-resistant upland cotton into a sensitive species of upland cotton. Controls of blank injections without DNA and same genotype DNA did not produce variations. However, verification by DNA analysis was not done. Subsequently, transformation frequency was verified by molecular analysis and ranged from 0.28 to 0.84 % in one instance and 1.13 to 1.21 % in another (Song et al. 2006). There has continued to be much interest in pursuing this avenue using marker genes amenable to physical selection and DNA analysis, (Reddy et. al., 2004; Mogali, et. al., 2007). A similar avenue was investigated by Burke et. al. (1999) who utilized cotton pollen germinated *in vitro* on *Agrobacterium* coated medium to pollinate greenhouse grown cotton flowers. Trolinder et al (1999b) utilized a needle-less injection method for delivery of *Agrobacterium* to the placental region of 1 DPA ovules.

Frequently, cotton transformation approaches follow those taken with other species. Recently Carlson et al (2007) announced the successful incorporation of *in vitro* assembled autonomous mini-chromosomes into maize. The mini-chromosomes were delivered by particle gun bombardment into embryogenic maize tissue. Results indicated that the autonomous mini-chromosomes were mitotically and meiotically maintained in transgenic progeny. This approach has the advantages of combining several traits on a single DNA fragment, arranging genes in a defined sequence for consistent gene expression, and providing an independent linkage group that can be rapidly introgressed into various germplasm. It is logical to assume that particle gun bombardment of any explant with mini-chromosomes tailored to that species, would be

applicable, given that the explant is capable of regenerating a whole plant whether through embryogenesis, organogenesis, or shoot apices.

Another option for delivery of mini-chromosomes to cotton might be incorporation directly into protoplasts, followed by regeneration of plants via somatic embryogenesis. Alternatively, Golovkin et al. (1993) were successful in obtaining transgenic plants by direct uptake of high molecular weight DNA into maize protoplasts. Willems et al. (2006), Chen et al. (1989), and Sun et al. (2005) reported successful regeneration of cotton via protoplast. Protoplasts fused with liposomes containing the mini-chromosomes or DNA, followed by induction of somatic embryogenesis and plant regeneration, will inherently suffer from the same drawbacks as from normal somatic embryogenesis, namely, genotype dependence and somaclonal variation. Nonetheless, the ability to stack multiple genes should make this an attractive model.

To address concerns of gene containment and low levels of transgene expression, gene silencing, positional and pleiotropic effects in transformed cotton, Kumar et al (2005) successfully obtained stable transformation of the cotton plastid genome. Transformation was effected by microprojectile bombardment and regeneration via somatic embryogenesis.

It is clear that many successful methods of cotton transformation exist. Still, at this time the most efficient method for commercial purposes remains somatic embryogenesis (Wilkins et al, 2004). Today many varieties have been identified as embryogenic and selections within those varieties may be made for highly embryogenic germplasm.

3.1 Transformation of Cotton – What to Engineer to do What?

Cotton can now be genetically engineered utilizing any one of the techniques outlined in the previous sections. The question then becomes how do we use this technology to understand and manipulate the cotton genome. From a crop production point of view, herbicide and insect resistant plants stood out as readily achievable and could provide the producer with less expensive, more effective, and more environmentally friendly insect and weed control. Today, the vast majority of the world's cotton acreage is planted with transgenic seed conferring insect and/or herbicide resistant traits based primarily on *Bacillus thuringiensis* (Perlak, 1990; Thomas et al, 1995) and glyphosate (Perlak, 2001). Appearance of resistant weed and insect species was anticipated and environmental steps to forestall such resistance were put into place (Deaton et al, 1995). Davis (2006) selected a naturally resistant cotton strain which clearly suggested that we would see resistant weeds quickly. As anticipated, resistance to glyphosate has been observed despite efforts to forestall such an event. Fortunately, transgenics resistant to other herbicides are becoming available (Rajaskeran et al, 1996, Stalker et al 1996, Trolinder et al., 2004). The use of transgenics with different mechanisms of action should allow producers to

successfully manage weed resistance. The release of herbicide and insect resistant transgenic cottons has played a powerful role in the breaking of a long held yield plateau in cotton production. Doors have been opened to a plethora of possible targets for improving cotton.

For the most part, attempts to alter plant gene expression through recombinant DNA techniques coupled with transformation/regeneration has dealt with single genes. Genetic engineering of multiple traits will prove more difficult. For cotton, fiber and seed comprise yield. Yield is dependent upon whole plant growth and development. The single most important factor in yield reduction is environmental stress. Traditionally, we have considered yield and stress responses to be very complex multigenic traits. Single genes known to be involved in plant response to environmental stress have been logical targets for expression in transgenic plants. Genes involved in response to the presence of reactive oxygen species generated during stressful conditions are thought to be control points for plant development (Bowler, et al, 1992; McKersie, et al, 1993; Allen, 1995; Gapper and Dolan, 2006). A number of genes involved in scavenging reactive oxygen species have been isolated and in some cases transferred to cotton. These include superoxide dismutases, glutathione reductase, and ascorbate peroxidase (Allen and Trolinder, 1995; Kornveyev et. al., 2001; Payton et. al., 2001; Logan et. al., 2003; Yan et.al., 2004). While there seems to be some positive effects on plant growth during stress due to altered expression of these genes, the results are still far from clear.

Alternative oxidase (AOX) is a branch point for metabolism and carbon partitioning during electron transport. AOX has been shown to have significant effects on reactive oxygen species associated with stress induced by low temperature, high temperature, and drought (Wagner and Moore, 1997, Moynihan et al, 1995). Lennon et al. (1997) demonstrated that tobacco mosaic virus has a similar effect on AOX activity as salicylic acid, a known inducer of AOX. Thus this enzyme seems to serve several regulatory functions during biotic and abiotic stress. Moynihan et al (1995) showed that AOX activity increases when cotton tissues are exposed to low or high temperature and growth is positively altered when the alternative oxidative pathway is induced. A cotton AOX gene has been sequenced (Li et al, 2007) and expression of the endogenous gene characterized. Hirut et al. (2007) transformed cotton with the well characterized AOX1 of tobacco to determine the effects of over expression of AOX on cotton growth and response to environmental stress, in particular low temperature. The spatial and temporal expression of two lines verified as single copy homozygotes was increased by 100 fold in stem, root and bolls when compared to the null line. It is anticipated that this over-expression will confer some advantage to cotton tissue experiencing biotic or abiotic stress.

Identifying critical genes associated with regulatory pathways is paramount for positively impacting growth under environmental stress through genome manipulation, whether traditional or transgenic breeding. To this end, close collaboration between biochemist, molecular biologists, physiologists, and plant breeders is essential.

Historically, producers have received little value from cotton seed, usually just enough to pay their ginning costs. If the cottonseed could be engineered to produce a more desirable high value product, or produce it more efficiently under environmental stress, the producer and the public would benefit. Foremost of concern in this chapter are seed oil composition and protein content. Literature concerning the fiber component of the cottonseed coat is voluminous and is covered elsewhere.

Cottonseed oil is 21% of the seed biomass and is the third largest source of vegetable oil. It is used primarily in the “chip” industry in the United States. Almost all chips (potato etc.) are fried in cotton seed oil because it produces a more tasteful product than other oils. However, it is not the most healthy oil because of the high level of palmitic acid, undesirable because of cholesterol raising properties. To alter cottonseed to a more desirable composition, Green et al. (2005) and Liu et al (2002) successfully increased the oleic component from a normal level of 18% to between 58.5% and 68.9% and decreased the palmitic level. This was accomplished by silencing two key fatty acid desaturase genes, ghSAD-1-encoding stearyl-acyl-carrier protein delta 9 – desaturase and ghFAD2-1-encoding oleoyl-phosphatidylecholine 6-desaturase. Hairpin RNA-encoding gene constructs were targeted against either ghSAD-1 or ghFAD2-1 and transformed into cotton. Huynh (2001) produced transgenic cotton plants with a slight increase in oleic acid content by silencing FatC encoding palmitoyl-ACP thioesterase which plays a major role in regulating palmitic acid of extraplastidial complex glycerolipids. Results indicate that altering palmitic acid by this method may have deleterious effects on membrane integrity. Another important question to be asked about any product with lower palmitic oil is “will chips fried with the new oil product pass the customer taste satisfaction test?”.

It would be desirable for cotton seed meal to provide a high protein cotton food product, particularly in third world countries where there is a growing requirement for food and 20 million farmers in Asia and Africa alone. Cottonseed contains 23% protein and global cotton cultivation produces in the range of 40 million metric tons of cottonseed annually. The production of edible cottonseed could provide the protein requirement of nearly 500 million people per year (Sunikumar et al 2006). Unfortunately, commercial cotton seed contains a toxic compound, gossypol, that renders it useful only to ruminant animals. Although there are wild species of cotton that do not contain gossypol and introgression has produced gossypol free cultivated cotton, the presence of gossypol in the plant itself offers some protection against pathogens and insects (Stipanovic et al 1977). Ideally, it would be desirable to have glanded plants and glandless seed.

Koshinsky et al. (1995) approached the problem with the knowledge that for every molecule in our environment, there is a bacterium somewhere that can degrade it. Therefore, bacteria were screened for their ability to degrade Gossypol. Gossypol degrading bacteria were identified and the associated genes isolated for transfer to cotton. To date, effective suppression has not been reported.

Townsend et al (2005) addressed the enzyme (+)- δ -cadinene synthase (CDNS), as the first step in the biosynthesis of sesquiterpenes, such as gossypol, that provide constitutive and inducible protection against pests and diseases. Of the five CDNS genes identified, anti-sense constructs for constitutive and seed specific expression of *cdn1-C4* were transformed into cotton. Gossypol levels were not reduced in seeds in either case, nor was induction of CDNS by *Verticillium dahliae* Kleb in stems. However induction was completely blocked in response to bacterial blight infection of cotyledons by constitutive expression, suggesting that different genes in the multigene family differ in their temporal and spatial regulation.

Researchers at Texas A&M University led gossypol research efforts for decades before there was the possibility of genetically engineering the cotton plant to produce gossypol only in the plant and not the seeds (Stipanovic et al 1986; Veech et al 1976). The very effective gene silencing system, RNAi, where the RNAi sequence had high homology to certain sequences contained in all cadinene genes in the multi-gene family, was driven by a highly specific seed promoter. The construct was transformed into cotton via standard *Agrobacterium*/somatic embryogenesis technology. The resulting transgenic plants were tested for the presence of gossypol in both seed and plant. There was an almost complete elimination of gossypol in the seed but no effect on the plant itself (Sunikumar et. al. 2006). Given the results of Townsend et al (2005), new questions must be asked about temporal regulation and pathogen resistance. Further, the elegant design of RNAi used in suppression of the cadinene multi-gene family can be taken into account for other multi-gene systems.

A third problem of cotton seed involves the seed producer. Attacking this problem involved reaching beyond simple traits. Historically, producers in some areas of the U.S. and other countries have not seen cottonseed as a profitable product and in an effort to conserve profits “catch” their own seed rather than purchase certified seed. While this practice in the short term saves the cost of seed, thereby providing a better profit margin, long term genetic erosion and poor seed quality due to environmental conditions can eliminate any savings and have overall consequences for yield. With the advent of new biotech products, there has been a rightful concern that companies marketing them cannot recover their research and development costs in areas and countries where saved seed is practiced, coupled with the lack of patent protection. In 1995 the use of site specific recombination to restrict the use of transgenic plants without proper authority was proposed by USDA and Delta and Pine Land Co (Oliver et. al, 1998, 1999a, 1999b). The announcement of this technology resulted in disagreement among producers, seed companies, and governments throughout the world. At the time the technology was referred to as “The Terminator” by the media. The technology details a method to control plant gene expression, and is based on the ability to induce a lethal condition in germinating transgenic seed. Originally, and for some years after the introduction of the concept, governments, environmental organization, and commercial organizations pledged to ban the technology from use. However, with the

passage of time and a clearer understanding of the potential value to agriculture, pledges to ban the technology are now disappearing.

The global research community renamed the technology GURT, Genes Under Restriction Technology. The original technology has been termed vGURTS and restrict the use of proprietary varietal technology as well as allows the production of hybrid cotton. Site specific recombination technology has many uses other than varietal protection. Modification of transgenic traits with this technology has been termed tGURTs. Bayley et al (1992) demonstrated the exchange of gene activity in transgenic plants catalyzed by the Cre-lox site-specific recombination system. Van Haaren et al, (1993) introduced the combination of DNA transposition and site specific recombination to modify chromosomes. Qin et. al. (1994) reported Cre mediated site specific recombination between plant chromosomes. Subsequently, Medberry et. al. (1995) reported intrachromosomal rearrangements generated by Cre-lox site specific recombination. Ow (2005) and Hodges and Lyznik (1996) utilized site specific recombination to target, replace, stack, or translocate genes. Chua et. al. (2004) used inducible site specific recombination for the activation and removal of transgenes in transgenic plants. Similarly, one may specifically target a transgene to a recombination site (Baszczynski, et al. 2008). Ow and Srivastava (1998) used site specific recombination to resolve tandemly repeated transgenes into one copy. Gateway Technologies (Earley et al 2006) has developed commercially available inducible expression vector constructs that provide tighter control than the GURT systems described by Oliver et. al. (1998, 1999, 1999). The 'double-lock' inducible system requires both heat shock and dexamethasone-control of cellular targeting of cyclization CRE recombinase in order to activate a promoter disrupted by a DNA fragment flanked by X-over P1 sites (lox sites) Specifically, heat shock is used to induce the expression of CRE recombinase fused to the hormone-binding domain of the rat glucocorticoid receptor. The resulting protein remains sequestered in the cytoplasm until dexamethasone treatment, which allows the protein to move into the nucleus, catalyzes the removal of the sequence blocking transcription by the 35S promoter, and thereby allows expression of the target gene.

Van Haaren et. al. (1993) and Medberry et. al. (1995) provided the nucleus for a model to produce transgenic gene knockout populations of cotton that will allow the systematic dissection of cotton chromosomes (Trolinder, N.L. T.A. Wilkins (2002). The ability to systematically move the ds element up and down a particular chromosome and excise the intervening regions will aid in the elucidation of gene function in relation to chromosomal architecture and genome sequence. A gene knock out population has been obtained with preliminary evidence for the movement of the ds element (Trolinder and Wilkins, 2002). At this time, the chromosomal location of each ds element has not been determined. With the completion of genome sequencing and the accompanying physical location of gene sequences, this population of transgenic plants, and others to be generated, will complement on-going efforts in dissecting and understanding the cotton genome.

Advances in site specific recombination will be critical to plant transformation in the post genomic era (Ow 2002). Clearly, advances in technology are rapidly moving forward our ability to understand and manipulate the cotton genome. We have attempted to demonstrate the steady progress over time in cotton genomics from simple tissue culture to regeneration to transformation of simple traits to very complex manipulations of the cotton genome. As we move forward, transformation and regeneration will remain critical components to understanding and altering the cotton genome in ways that will benefit the worldwide community.

References

- Agrawal, D.C., A.D. Banerjee, R.R. Kulala, A.B. Dhage, A.B., A.V. Kulkarni, M. Nalawade, S.L. Hazra, K.V. Krishnamurty, K.V. (1997). *In vitro* induction of multiple shoots and plant regeneration in cotton (*Gossypium hirsutum* L). *Plant Cell Reports* 16(9) : 647–652
- Allen, R.D. (1995). Dissection of oxidative stress tolerance using transgenic plants. *Plant Physiol.* 107: 1049–1054.
- Allen, R.D. and N.L. Trolinder (1995). Expression of superoxide dismutase in transgenic plants leads to increased stress tolerance. *Proceedings Beltwide cotton Conference*, Vol. 2: 1136–1137. National Cotton Council, TN.
- Armstrong, Toni A., D. L. De Boer, (2004). Method for regeneration of cotton. U.S. patent application 20040087030
- Bajaj, Y.P.S. and M. S. Gill (1985). In vitro induction of genetic variability in cotton (*Gossypium* spp.). *Theor. Appl. Genet.* 70:363–368.
- Bajaj, Y.P.S. and M.S. Gill (1986) Micropropagation and germplasm preservation of cotton (*Gossypium* spp.) through shoot tip and meristem culture. *Ind. J. Exp. Bot* 24:581–583.
- Barrow, J. R. (1986) The conditions required to isolate and maintain viable cotton (*Gossypium hirsutum* L.) microspores. *Plant Cell Reports* 5(6)
- Bayley, C.C.; Morgan, M.; Dale, E.C., Ow, D.W. 1992. Exchange of gene activity in transgenic plants catalyzed by the Cre-lox site-specific recombination system. *Plant Molecular Biology* 18: 353–362
- Bayliss, M.W. (1973) Origin of chromosome number variation in cultured plant cells. *Nature* 246: 529–530.
- Beasley, C.A. (1971). In vitro culture of fertilized cotton ovules. *Bioscience* 21 906–907.
- Beasley, C.A. and I.P. Ting (1973). The effects of plant growth substances on in vitro fiber development from fertilized cotton ovules. *Amer J. Bot* 60:130–139.
- Baszczyński, C. L., B.A. Bowen, D.J. Peterson, L. Tagliani (2008) Compositions and methods for the targeted insertion of a nucleotide sequence of interest into the genome of a plant. U.S. patent 7,361,508.
- Bowler, C., M. Van Montagu, D. Inze (1992). Superoxide dismutase and stress tolerance. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 43:83–116.
- Burke, John J; M.J. Oliver, J.P. Velten (1998). Pollen based transformation system using solid media U. S. patent 5,929,300.
- Cao, Jing-Lin; X. Zhang, S.X. Jin, X.Y. Yang, H.G. Zhu, L-L Fu, (2008). An efficient culture system for synchronization control of somatic embryogenesis in cotton (*Gossypium hirsutum* L). *Acta Agronomica Sinica* 34(2) 224–231.
- Carlson, S.R., G.W. Rudgers, H. Zieler, J.M. Mach, S. Luo, E. Grunden, C. Krol, G.P. Copenhaver, D. Preuss (2007). Meiotic Transmission of an In Vitro- Assembled Autonomous Maize Minichromosome. *PloS Genet* 3(10): e 179. doi:10.1371/journal.pgen.0030179.

- Chen, Z. X., S. J. Li, J. X. Yue, G. L. Jiao, and S. X. Liu (1989). Plantlet regeneration from protoplasts isolated from an embryogenic suspension culture of cotton (*Gossypium hirsutum* L.). *Acta Botanica Sinica* 31,966–9.
- Chappell, J., and J.R. Mauney (1967). Culture of the apical meristem of *Gossypium hirsutum* in vitro. *Phyton* 24: 93–100.
- Chlan, C. A., J. Lin, J. W. Cary, and T. E. Cleveland, (1995). A procedure for biolistic transformation and regeneration of transgenic cotton from meristematic tissue. *Plant Mol. Biol. Rep.* 13, 31–7.
- Chua, N.H., J. Zuo, S. G, Moller (2004). Inducible site specific recombination for the activation and removal of transgenes in transgenic plants. U.S. Patent 6,723,896.
- Cousins, Y. L., B. R. Lyon, and D. J. Llewellyn (1991). Transformation of an Australian cotton cultivar: prospects for cotton improvement through genetic engineering. *Aust. J. Plant Physiol.* 18,481–94.
- Davidonis, G. H., and R. H. Hamilton (1983). Plant regeneration from callus tissue of *Gossypium hirsutum* L. *Plant Sci. Lett.* 32,89–93.
- Davis, B. 2006. Genetically controlled herbicide resistance in cotton plants in the absence of genetic engineering. US Patent 7,074,987.
- Deaton, W.R. (1995). Managing for resistance to the Bollgard gene. IN: Proceedings of the Beltwide Cotton Conference, Vol 2: 758. National Cotton Council. Memphis, TN.
- Earley, K.W., J.R. Hang, O. Pontes, K. Oppen, T. Juehne, F. Song, C.S. Pikaard (2006). Gateway compatible vectors for plant functional genomics & proteomics. *The Plant Journal* 45: 616–629.
- Finer, J. J. and Smith, R. H. 1984. Initiation of callus and somatic embryos from explants of mature cotton (*Gossypium klotzschianum* Anderss). *Plant Cell Reports* 3,41–43.
- Finer, J., 1988. Plant regeneration from somatic embryogenic suspension cultures of cotton (*Gossypium hirsutum* L.). *Plant Cell Rep.* 7, 399–402.
- Finer, J. J. and McMullen, M. D. 1990. Transformation of cotton (*Gossypium hirsutum* L.) via particle bombardment. *Plant Cell Reports* 8,586–9.
- Firoozabady, E. Plant regeneration via somatic embryogenesis in many cultivars of cotton (*Gossypium hirsutum* L. (1993). *In Vitro Cellular and Developmental Biology – Plant* 29 (3).
- Firoozabady, E., Deboer, D., Merlo, D., Halk, E., Amerson, L., Rashka, K. and Murray, E. 1987. Transformation of cotton (*Gossypium hirsutum* L.) by *Agrobacterium tumefaciens* and regeneration of transgenic plants. *Plant Mol. Biol.* 10: 105–16.
- Gapper, C. and L. Dolan (2006). Control of Plant Development by Reactive Oxygen Species. *Plant Physiol.* 141(2):341–345.
- Gawel, N. J. and Robacker, C. 1990. Genetic control of somatic embryogenesis in cotton petiole callus cultures. *Euphytica* 49,249–53.
- Golovkin, M.V., M. Abraham, S. Morocz, S., Bottka, A. Feher, and D. Dudits, (1993). Production of transgenic embryogenic plants by direct DNA uptake into maize protoplasts. *Plant Sci.* 90:41–52.
- Gould, J., S. Banister, O. Hasegawa, M. Fahima, R.H. Smith, (1991). Regeneration of *Gossypium hirsutum* and *G. barbadense* from shoot apex tissues for transformation. *Plant Cell Reports* 10,12–6.
- Green, A., S. Singh, Q. Liu, (2005). Method of modifying the content of cotton seed oil. U.S. patent 6,974,898.
- Hemphill, J.K., C. G. Maier, K.D. Chapman, (1998). Rapid in vitro plant regeneration of cotton (*Gossypium hirsutum* L.). *Plant Cell Rep.* 17, 273–278.
- Hodges, T.K., L.A. Lyznik (1999). Controlled modification of eukaryotic genomes. U.S. patent 5,910,415.
- Hood, E.E., S.B. Gelvin, L.S. Melchers, A. Hoekema (1993). New *Agrobacterium* helper plasmid for gene transfer to plants. *Transgenic Res.* 2: 208–218.
- Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. G. and Fraley, R. T.(1985). A simple and general method for transferring genes into plants. *Science* 227: 1229–1231.

- Huynh, T.T. (2001) Palmitol-Acyl Carrier Protein Thioesterase in Cotton (*Gossypium hirsutum* L.): Biochemical and Molecular Characterization of a Major Mechanism for the Regulation of Palmitic Acid Content. Dissertation. University of North Texas
- Huzera, S; A. V. Kulkarni, S. M. Naluwade, A.K. Bienerjee, D.C. Agrawal, K.V. Krishnamurthy. (2006) Multiple shoot regeneration in cotton: Influence of explants, genotypes and culture vessels on sprouting and proliferation of pre-existing meristems of cotton (*Gossypium hirsutum* L and *Gossypium Arboreum* L). IVC and DB 36 (6) 505–510.
- Jiang, B. (2004). Optimization of *Agrobacterium* mediated cotton transformation using shoot apices explants and quantitative trait loci analysis of yield and yield component traits in upland cotton (*Gossypium hirsutum* L). A Dissertation. Louisiana State University.
- Jain, S.M.; B.S. Ahloorolia, D. S. Brar, (1998). Somaclonal variation and induced mutations in crop improvement. Ed. Kluwer Academic Press, The Netherlands.
- Jim, S., X. Zhang, Y. Nie, X. Guo, S. Liang, H. Zhu. (2006). Identification of a novel elite genotype for *in vitro* culture and genetic transformation of cotton. *Biologia Plantarum* 50 (4)
- Kebede, D. Tripathy, S. Hwang, N. Trolinder, R. Wright (2007). Alternative respiration during cotton growth & development. International Cotton Research Conference 2007, Lubbock, Texas.
- Kim, J.K., B.A. Triplett (2001). Cotton fiber growth *In Planta* and *In Vitro*: Models for plant cell elongation and cell wall biogenesis. *Plant Physiol* 127: 1361–1366.
- Klein, R.M., E.D. Wolf, R. Wu., J.C. Sanford (1987). High-velocity micro-projectiles for delivering nucleic acids into living cells. *Nature*, 327: 70–73.
- Korneyev, D. B., A. Logan P., Payton, R. D. Allen, A. S. Holaday (2001) Enhanced photochemical light utilization and decreased chilling-induced photoinhibition of photosystem II in cotton overexpressing genes encoding chloroplast-targeted antioxidant enzymes. *Physiologia Plantarum* 113: 323–331
- Koshinsky, H.A., H. K. Liao, D.W. Ow (1995). Progress in screening micro-organisms for gossypol degrading ability. Biochemistry of cotton (Proceedings of the biochemistry of cotton workshop) Galveston, TX. Cotton Incorporated pp 19-22.
- Kosegi, B. D, J.R. Beringer, A. Palta, A. Mehra, J.F..Petoline, R. Ram, (2007). Whisker-mediated transformation of embryogenic cotton suspension cultures. US PATENT 7,166,768.
- Kumar, S., P. Sharma, D. Pentel. A genetic approach to *in vitro* regeneration of non-regenerating cotton (*Gossypium hirsutum* L.) cultivars. (1998). *Plant Cell Rep.* 18: 59–63.
- Kumar, S. , A. Dhingra, H. Daniell (2005). Stable transformation of the cotton plastid genome and maternal inheritance of transgenes. *Plant Molecular Biology* 56: 203–216.
- Kumria, R., V.G. Sunnichan, D.K. Das, S.K Gupta, V.S. Reddy, R.K Bhatnagar, S. Leelavathi (2003). High-frequency somatic embryo production and maturation into normal plants in cotton (*Gossypium hirsutum*) through metabolic stress. *Plant Cell Rep.* 21:635–639.
- Lennon, A.M. U.H.Neuenschwander, M. Ribas-Carbo, L. Giles, J.A. Ryals, J.N. Siedow (1997). The effects of salicylic acid and tobacco mosaic virus infection on the alternative oxidase of tobacco. *Plant Physiol* 115: 783–791.
- Li, L., Y. Zhang, M. Wang, Y. Zhang, X. Wu, X. Guo. (2007). Molecular cloning and expression characteristics of alternative oxidase gene of cotton. *Molecular Biology Reports*: on line 1573-4928
- Li, R., D.M. Stelly and N.L. Trolinder (1989). Cytogenetic abnormalities in cotton (*Gossypium hirsutum* L.) cell cultures. *Genome* 32: 1128–1134.
- Liu, Q. S.P. Singh, and A. G. Green (2002). High-Stearic and High-Oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Phys.* 129: 1732–1743
- Logan, B.A. G. Monteiro, D. Korneyev, P. Payton, R. D. Allen and A.S..Holaday (2003). Transgenic overproduction of glutathione reductase does not protect cotton, *Gossypium*

- hirsutum* (Malvaceae), from photoinhibition during growth under chilling conditions. American Journal of Botany 90:1400–1403.
- Luo, J-H and Gould, J.H. (1999). *In vitro* shoot-tip grafting improves recovery of cotton plants from culture. Plant Cell Tissue and Organ Culture 57, 211–213.
- Maxwell, D.P., Y. Wang, L. McIntosh (1999). The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. Proc Natl Acad Sci USA 96:8271–8276.
- McCabe, D. E. and Martinell, B. J. (1993). Transformation of elite cotton cultivars via particle bombardment of meristems. Bio/technol. 11,596–8.
- McKersie, B.D., Y. Chen, M deBeus, S.R. Bowley, C Bowler, D. Inze, K. D'Halluin, J. Botterman. (1993). Superoxide dismutase enhances tolerance of freezing stress in transgenic alfalfa (*Medicago sativa* L.). Plant Physiol. 103:1155–1163.
- Medberry, S.L., E. Dale, M. Qin, D. W. Ow. (1995). Intra-chromosomal rearrangements generated by Cre/lox site specific recombination. Nucleic Acids Research 23 - 485–490.
- Mishra, R., H-Y Wang, N. Yadav, and Wilkins, T.A. (2003). Development of highly regenerable elite Acala cotton (*Gossypium hirsutum* L.) A step towards genotype-independent regeneration. Plant Cell Tissue Organ Culture 73, 21–39.
- Mogali, S.C., B.M. Khadi, I.S. Katageri (2007). Pollen tube pathway mediated genetic transformation studies in cotton (*Gossypium hirsutum* L.). International Cotton Research Conf., Lubbock, Texas.
- Oliver, M.J., J.E. Quisenberry, N.L. Trolinder, D.L. Keim (1998) Control of plant gene expression U.S. patent 7,723,765
- Oliver, M.J., J. E. Quisenberry, N.L. Trolinder, D. L. Keim (1999a). Control of Plant Gene Expression. U.S. Patent 5,979,441.
- Oliver, M.J., J.E. Quisenberry, N.L. Trolinder, D.L. Keim (1999b). Control of plant gene expression. U.S. patent 5,925,808
- Payton P., R. Webb, D. Korniyev, R. Allen, A. S. Holaday (2001). Protecting cotton photosynthesis during moderate chilling at high light intensity by increasing chloroplastic anti-oxidant enzyme activity. Journal of Experimental Botany 52: 2345–2354
- Perlak, F. J., R. W. Deaton, T. A. Armstrong, T.A. Fuchs, T. S.R. Sims, J. T. Greenplate. and D.A. Fischhoff. (1990). Insect resistant cotton plants. Bio/technol. 8,939–943.
- Perlak, Fredrick J. (2001). Development and commercial use of Bollgard R cotton in the USA – Smith, R. H., H. J.Price, J.B.Thaxton(1977). Defined conditions for the initiation and growth of cotton callus in vitro. I. *Gossypium arboreum*. In Vitro 13(5): 329–334.
- Price, H. J., R.H. Smith (1979). Somatic embryogenesis in suspension cultures of *Gossypium klotzschianum* Anderss. Planta 145,305–6.
- Qin, M.; D. Baley, J. Stockston, D.W. Ow, (1994). Cre recombinase mediated site-specific recombination between plant chromosomes. Proceedings National Academy Science USA 91: 1706–1710.
- Quma, J.P.; Young, M.M.; and Reichert, N.A. (2004). Rooting of *in vitro* regenerated cotton (*Gossypium hirsutum* L) is influenced by genotype, medium composition, explants type and age. African Journal of Biotechnology 3 (6): 313–318.
- Rajasekaran, K., Grula, J. W., Hudspeth, R. L., Pofelis, S. and D. M. Anderson, (1996). Herbicide-resistant Acala and Coker cottons transformed with a native gene encoding mutant forms of acetohydroxyacid synthase. Mol. Breeding 2: 307–19.
- Rangan, T. S.; D. M. Anderson, M. David (1998). Method for producing somaclonal variant cotton plants. U.S. Patent 5,834,292
- Rangan, T. S. D.M. Anderson, K. Rajasekaran (1999). Cotton somaclonal variants with increased resistance to fungal pathogens; U.S Patent 5,859,321.
- Rangan, T. S., K. Rajasekaen (1997). Regeneration of cotton plants in suspension culture. U. S. patent 5,695,999.
- Rangan, T.S. and T. Zavala (1984) . Somatic embryogenesis in tissue culture of *Gossypium hirsutum* L.). In Vitro 20: 256.

- Rangan, T. S., D. M. Anderson, K. Rajasekaran, J.W. Grula, R. L. Hudspeth, R. L. Yenofsky, (2003). Transformation of cotton plants. U.S. Patent 6,724,344.
- Reddy, N.S., J.Y. Zhu, J.Y. Rong, T.X. Hong (2004) Pollen tube pathway mediated genetic transformation of cotton. P 356-360. In "Strategies for sustainable cotton production – a global vision" Proc. Intern. Symp., 23-25 Nov. 2004, UAS, DWR, KRK.
- Reichert, N. A. T. Lim, M. Young, (2002). Method for transformation of cotton and organogenic regeneration. U.S. Patent 6,479,287.
- Reynaerts, A. , A. De Sonville (2002). Method for *Agrobacterium* mediated transformation of cotton . U.S. Patent 6,483,013.
- Shang, X. M., N. Trolinder (1991). Buffer Capacity of cotton cells and effects of extracellular pH on growth and somatic embryogenesis in cotton cell suspensions. *In Vitro Cellular and Developmental Biology – Plant* 27 (3): 147–152.
- Shoemaker, R.C., L.J. Couche, and D.W. Galbraith (1986). Characterization of somatic embryogenesis and plant regeneration in cotton (*Gossypium hirsutum* L.) *Plant Cell Reports* 3: 178–181.
- Song X. , Y. Gu * , G. Qin 2007 Application of a transformation method via the pollen-tube pathway in agriculture molecular breeding. *Life Science Journal*;4(1):77–79 J
- Stalker, D.M., J.A. Kiser, G. Baldwin, B. Coulombe, C.M. Houck (1996). Cotton weed control using the BXN system. In: *Herbicide-resistant crops: agricultural, environmental, economic, regulatory, and technical Aspects*, pp 93–105. Duke, S.O. Ed., Lewis Publishers, New York.
- Stelly, D.M., D. W. Altman, R.J. Kohel, T.S. Rangan, and E. Commiskey (1989). Cyatogenetic abnormalities of cotton somaclones from callus cultures. *Genome* 32, 762–770.
- Stewart, J. McD, C.L. Hsu (1977). *In ovulo* embryo culture and seedling development of cotton (*Gossypium hirsutum* L.). *Planta* 137:113–117.
- Stipanovic, R.D., A. Stossel, J.B Stothers, D.W. Altman, A.A. Bell, P Heinsteins (1986). The stereochemistry of the biosynthetic precursor of gossypol. *J Chem Soc Chem Comm* 2: 100–102.
- Stipanovic, R.D., A.A. Bell, M.J. Lukefahr (1977) Natural insecticides from cotton (*Gossypium*). In PA Hedin, ed, *Host Plant Resistance to Pests*, Vol 62. American Chemical Society Symposium Series, Washington, DC, 197–214.
- Strickland, S.G. (1998) Cotton transformation. U.S. Patent 5,846,797.
- Sun, Y. X. Zhang, C. Huang, ; Y. Nie, X. Guo (2005). Plant regeneration via somatic embryogenesis from protoplasts of six explants in Coker 201 (*Gossypium hirsutum*) *Plant Cell Tissue and Organ Culture* 82 (3): 309–315
- Sun, Y. (2006). Somatic embryogenesis and plant regeneration from different wild diploid cotton (*Gossypium* species). *Plant Cell Reports* 25 (4).
- Sunilkumar, G., L.M. Campbell, L. Puckhaber, R.D. Stipanovic, and K.S. Rathore (2006). Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *PNAS* 103(480): 18054–18059.
- Thomas, J. C., D. G. Adams, V. D. Keppene, C.C. Wasmann, J. K. Brown, M. R. Kanost, H. J. Bohnert (1995). Protease inhibitors of *Manduca sexta* expressed in transgenic cotton. *Plant Cell Rep.* 14:758–62.
- Townsend, B.J., A. Poole, C.J. Blake, D.J. Llewellyn (2005). Anti-sense Suppression of a (+) - &- Cadinene Synthase gene in cotton prevents the induction of this defense response gene during bacterial blight infection but not its constitutive expression. *Plant Physiol.* 138:516–528.
- Trolinder, N. L .(1985). Somatic embryogenesis and plant regeneration in cotton (*Gossypium hirsutum* L.). A Dissertation in Biology (Dec., 1985) Texas Tech Univ.
- Trolinder, N. L. (1987) . Genotype specificity of the somatic embryogenesis response in cotton. *Plant Cell Reports* 8,133–6.
- Trolinder, N. L. and J. R. Goodin, (1987). Somatic embryogenesis and plant regeneration in cotton (*Gossypium hirsutum* L.). *Plant Cell Reports* 6: 231–4.

- Trolinder N. L., J.D.Berlin, Goodin JR (1987) Differentiation of cotton fibers from single cells in suspension culture. *In Vitro Cell Develop Biol* 23:789–794
- Trolinder, N. L., J. R. Goodin (1988a). Somatic embryogenesis and regeneration in cotton. I. Effects of source of explant and hormone regime. *Plant Cell Tissue Organ Culture* 12: 31–42.
- Trolinder, N. L., J. R. Goodin (1988b). Somatic embryogenesis and regeneration in cotton. II. Requirements for embryo development and plant regeneration. *Plant Cell Tissue Organ Culture* 12: 43–53
- Trolinder, N. L., X. Shang (1991). In vitro selection and regeneration of cotton resistant to high temperature stress. *Plant Cell Reports Vol* 10 (9): 448–452..
- Trolinder, N.L., J. G Dever, and L. Koonce (1999a) Transformation and regeneration of fertile cotton plants. U.S. Patent 5,986,181.
- Trolinder, N.L. (1999b) In planta method for production of transgenic plants. U.S. Patent 5,994,624.
- Trolinder, N.L., T.A. Wilkins (2002). Creation of a gene knockout population of cotton, International Cotton Research Initiative Workshop, Nanjing, China S561.
- Trolinder, L., J. Gwyn, M. Debeuckeleer (2004). Herbicide tolerant plants having event EE-GH1. U.S. Patent 6,818,807.
- Trolinder, N. L.; L.K. Koonce. J.K Dever. (2006). Methods for producing transgenic cotton plants using chilled apical shoot tips. US Patent 7,122,722.
- Tuli, R.; A. K. Srivastana, S. K. Gupta, (2001). Tissue Culture process for producing a large number of viable cotton plants in vitro. U.S. Patent 6,242,257
- Ulian, E.C., R.H. Smith, J.H. Gould, and T. D. McKnight. (1988). Transformation of plants via the shoot apex. *In Vitro Cell & Devel. Bio.* 24:951–954.
- Umbeck, P. F. (1991). Genetic engineering of cotton plants and lines. US Patent No. 5,004,863.
- Umbeck, P. F. (1992). Genetic engineering of cotton plants and lines. US Patent No. 5,159,135
- Umbeck, P. F, Johnson, G., Barton, K. and Swain, W. (1987). Genetically transformed cotton (*Gossypium hirsutum* L.) plants. *Bio/technol.* 5,263–6.
- Van Haaren, M. J.J.; Ow, D.W. (1993). Prospects of applying a combination of DNA transposition and site specific recombination in plants. *Plant Molecular Biology* 23: 525–533.
- Veech, J.A., R.D. Stipanovic, A.A. Bell (1976). Peroxidative conversion of hemigossypol to gossypol: a revised structure for isohemigossypol. *J Chem Soc Chem Comm*: 144–145
- Wagner, A.M., A.L. Moore (1997) Structure and function of the plant alternative oxidase: its putative role in the oxygen defence mechanism. *Biosci Rep* 17:319–333.
- Wang, Y., X., Z. Ma, G. Zhang, G. Han. (2006). Somatic embryogenesis and plant regeneration from two recalcitrant genotypes of *Gossypium hirsutum* L. *Chinas Agricultural Science* 05.
- Wilkins, T. A., R. Mishra, and N.L. Trolinder (2004). Agrobacterium-mediated transformation and regeneration of cotton. *Food, Agriculture & Environ* 2(1):179–187.
- Willems, K, R. Swennen (2006). Protoplast to plant regeneration in cotton (*Gossypium hirsutum* L cv Coker 312) using feeder layers. *Plant Cell Reports* 13(3–4) 208–211.
- Yan, J., C. He, J. Wang, Z. Mao, S.A. Holaday, R. D. Allen and H. Zhang. (2004). Over-expression of the *Arabidopsis* 14–3-3 Protein GF14 in Cotton Leads to a “Stay-Green” Phenotype and Improves Stress Tolerance under Moderate Drought Conditions. *Plant and Cell Physiology* 45(8):1007–1014.
- Zeng, F., X. Zhang, L. Zhu, X. Guo, Y. Nie (2006). Isolation and characterization of genes associated to cotton somatic embryogenesis by suppression subtractive hybridization and macroarray. *Plant Molecular Biology Vol* 60(2) 167–183.
- Zhong, H., E. Boudreau, S. Rouse, E. Dunder, W.. Gu, Y. Chang, (2005). Methods for stable transformation of plants. U. S. patent 6,858,777.

- Zhang, B. R. Feng, F. Lin, Q. Wang, (2001). High frequency somatic embryogenesis and plant regeneration of an elite Chinese cotton variety. *Bot. Bull Acad Sin* 42: 9–16.
- Zhou, G.-Y., Weng, J., Zeng, Y.-S., Huang, J.-G., Qian, S.-Y. and Liu, G.-L. 1983. Introduction of exogenous DNA into cotton embryos. *Methods in Enzymology* 101,433–81.

Mutagenesis Systems for Genetic Analysis of *Gossypium*

Dick Auld, Ginger G. Light, Mohamed Fokar, Efrem Bechere,
Randy D. Allen

Abstract The recent evolution of tetraploid cotton combined with intensive selection of cultivated cottons has reduced the genetic diversity of cotton. This lack of allelic diversity hampers efforts to improve the agronomic traits of cotton and limits the application of molecular genetic tools for improvement of cotton germplasm. The lack of genetic resources also reduces our ability to understand the molecular mechanisms that regulate cotton growth and development and its responses to environmental stresses and pathogens. Use of a variety of chemical mutagens and ionizing radiation can be used to effectively increase the frequency of mutant alleles in *Gossypium* species. While application of insertional mutagenesis methodologies that require high-throughput plant transformation procedures is not feasible, evaluation of various transposon-based mutagenesis systems is underway. TILLING technology, which uses a combination of mutagenesis and high-throughput molecular screening methods for reverse genetics is also being developed.

1 Introduction

World wide the production of cotton suffers from competition with synthetic fibers in the manufacture of yarns and textiles because of both poor fiber quality characteristics and the high cost of production. Genetic enhancement of the length, strength and uniformity of cotton fiber would improve the competitiveness of cotton with synthetic fibers in the production of superior quality textiles (National Cotton Council, 2005). Genetic enhancement of the pest and environmental stress tolerance of cotton would improve the lint yield and ability of cotton to compete with synthetic fibers on a cost competitive basis. Unfortunately, the commercially grown cotton species have very limited genetic variability that can be utilized in the development of improved cultivars (May, et al., 1995).

D. Auld (✉)

Department of Plant and Soil Sciences, Texas Tech University
e-mail: dick.auld@ttu.edu

Several evolutionary bottlenecks have severely reduced the genetic variability available in modern cultivated cotton (Hutchinson et al., 1947 and Wendel, 1989). In addition, several decades of intensive breeding for improved cotton fiber characteristics has further narrowed the gene pool available for cotton improvement (Bowman, et al., 1996). Mutagenesis in combination with the application of modern genetic tools has the potential to rapidly increase the genetic variability of specifically targeted loci in cotton for both environmental and biotic stress tolerance as well as improve fiber characteristics.

2 Chemical and Radiation-Mediated Mutagenesis

The use of induced mutation has been highly successful in increasing genetic variability in economically important traits of many major crops such as rapeseed (Auld, et al., 1992, Tonnemaker, et al., 1992, Spasibionek, 2006), sugarbeets (Hohmann, et al., 2005) and rice (Wu, et al., 2005). Chemical mutagenesis was also a critical tool in early genetic mapping of the *Arabidopsis* genome (James and Dooner, 1990). However, mutation has historically played a less significant role in the improvement of cotton (Auld, et al., 1998).

Radiation mutagenesis has been shown to be an effective tool in the creation of a wide range of phenotypic variants in tetraploid (*G. hirsutum*) populations. Mutants were identified with improved earliness (using gamma rays as a mutagen) (Kandhro, et al., 2002; Xanthopoulos and Kechagia, 2001 and 2003); enhanced phosphorous uptake (radiophosphorous) (Nazirov, et al., 1981); improved drought tolerance (radiophosphorous) (Nazirov and Satipov, 1979); increased lint percent (X rays) (Cornellius, et al., 1970); as well as improvements in both earliness and lint yield (gamma rays) (Shamsuzzaman, et al., 2003). Radiation has also been used to create variation in the electron transport and the Hill Reactions (gamma rays) (Yakubova, et al., 1975); photo-period insensitivity (gamma rays) (Raut, et al., 1971); nonspecific esterases (gamma ray) (Gulin, et al., 1985); fiber properties (X rays) (Mehetre and Thombre, 1983); cytoplasmic sterility (gamma rays) (Negmatov, et al., 1975); and pollen size (gamma rays) (Savaskan, 2002). However, five cycles of seed irradiation at only 10kr of gamma rays failed to induce high levels of variation in both lint yield and fiber quality characteristics (Fotiadis and Miller, 1973).

Chemical mutagenesis has identified mutants in both *G. barbadense* and *G. hirsutum*. Huessin, et al., 1982, working with *G. barbadense* identified glandless cotton mutants using sodium azide as a mutagenic agent. Mutation frequency in *G. hirsutum* was shown to have a linear response to increasing concentrations of sodium azide (Larik, et al., 1983). Other researchers using *G. hirsutum* were able to identify resistance to Verticillium wilt (dimethyl sulfate) (Gaibullaev, et al., 1975), and improved fiber characteristics (ethylmethane sulfonate) (Auld, et al., 2000). Shattuck and Katterman (1982) found

that cotyledonary tissue of *G. barbadense* showed an unscheduled and enhanced DNA synthesis after treatment with ethylmethane sulfonate.

Colchicine has also been used to create genetic variation in lint yield, lint percent and fiber quality (Luckett, 1989); partially naked seed (Salanki and Parameswarappa, 1968); and parthenogenesis (Shi-Qi, et al., 1991). Katterman, 1973, found a decrease in the 5-methyl cytosine content in the DNA of colchicine and spontaneously induced polyhaploids of both *G. hirsutum* and *G. barbadense* but did not report any morphological or agronomic mutants. The treatment of cotton seed with colchicine in combination with gamma radiation (Sanamyan and Rakhmatullina, 2003) or with gamma radiation alone (Sheidai, et al., 2002) resulted in increased chromosomal aberrations.

Historically, nearly all of the mutant populations developed in cotton have been screened phenotypically for one or more specific trait(s) using a forward genetics approach to plant improvement (Herring, et al., 2004). Individual plants selected from the initial screening are then often progeny tested to ensure they are homozygous and breed true for the selected trait (Auld, et al., 1998). While this process has been moderately effective in increasing the genetic variation in cotton, it has been limited to enhancement of simply inherited traits which can be selected directly from the phenotype. Screening populations for specific enzymes using techniques such as TILLING (Targeting Induced Local Lesion IN Genomes) could lead to advances in both functional genetics and conventional genetic improvement of cotton.

3 One Example of the Potential for Identifying New Variation in Mutant Cotton Populations

In the spring of 2002, two kg of Fiber Max 958 (*G. hirsutum*) seed were exposed to a 3% v/v of Ethylmethane Sulfonate (EMS) at Texas Tech University. Untreated cotton seed (M_0 generation) were imbibed for 16 to 20 hours in aerated distilled water. The seed were rinsed and placed in aerated distilled water and the EMS treatment was pipetted into the solution where it mixed with the seed for two hours to produce the M_1 generation. The seeds were then removed, rinsed several times, and hand planted in the field at Lubbock, TX. In the fall of 2002, the M_2 generation was obtained by randomly harvesting one boll per plant. The M_2 generation was increased to the M_3 in 2004, the M_4 in 2005, and the M_5 in 2006 using this same process. In the fall of 2006, seed cotton was harvested from 500 individual M_5 plants and the fiber was screened for fiber quality using HVI analyses at the ITC - Fiber and Biopolymer Research Institute.

The 3% v/v EMS treatment reduced seed viability of the M_1 to less than 50% but did not appear to impact viability of further generations. Fiber from the 435 individual M_6 plants of FM 958 had significant variation for all five HVI Fiber quality characteristics measured (Table 1) but this variation was most pronounced for fiber length (Figure 1). Based on these results, 18 M_6

Table 1 HVI cotton fiber properties of 18 M₆ plants selected for short fiber length and 18 M₆ plants selected for long fiber length at Lubbock, TX in 2006

Sample #	MIC	LENGTH	UNIF.	STRENGTH	ELON.
		–in–	–%–	–g/tex–	
Short Fiber Length					
162	5.1	0.97	84.6	26.6	7.9
371	4.9	0.99	82.9	26.0	6.5
483	5.1	0.99	82.8	26.2	6.4
200	4.6	1.00	81.7	28.3	5.9
332	5.1	1.02	83.4	28.5	6.6
275	5.2	1.03	83.9	30.8	4.8
288	5.5	1.03	83.8	25.6	6.3
179	3.9	1.04	81.9	28.8	7.4
467	4.1	1.04	81.3	26.5	5.6
386	4.5	1.04	80.9	28.4	6.0
211	5.2	1.04	81.6	24.9	8.3
249	3.6	1.05	80.7	27.8	7.6
346	3.7	1.05	82.6	26.3	4.2
207	4.0	1.05	84.0	29.0	8.0
458	4.9	1.05	81.9	27.7	7.8
346	3.7	1.05	82.6	26.3	4.2
207	4.0	1.05	84.0	29.0	8.0
458	4.9	1.05	81.9	27.7	7.8
Long Fiber Length					
180	4.1	1.25	82.8	31.6	5.1
245	2.9	1.26	80.9	31.4	5.6
453	2.9	1.26	78.3	29.8	6.0
308	3.0	1.26	79.0	31.0	5.5
353	3.9	1.26	83.2	30.1	5.3
429	4.1	1.26	82.0	28.8	4.4
352	4.2	1.26	81.7	27.7	5.1
317	3.2	1.27	79.8	25.0	5.3
47	4.4	1.27	82.4	29.6	5.6
260	4.7	1.27	82.9	28.7	6.1
70	3.5	1.28	82.3	30.7	6.0
78	3.6	1.28	80.9	30.5	5.1
16	4.2	1.28	83.1	28.2	4.8
112	3.0	1.29	84.7	35.0	5.1
72	3.4	1.29	81.9	32.4	6.1
449	3.3	1.30	86.4	35.1	6.0
198	3.8	1.30	84.1	26.3	7.4
203	4.8	1.31	83.9	33.4	2.9
Average:	4.19	1.15	82.51	28.43	6.04
Std:	0.59	0.06	1.53	2.04	0.95
CV%	14.20	4.85	1.86	7.18	15.76
Highest:	5.50	1.31	86.40	35.10	9.40
Lowest:	2.30	0.97	77.40	23.50	2.90
Number samples	435	435	435	435	435

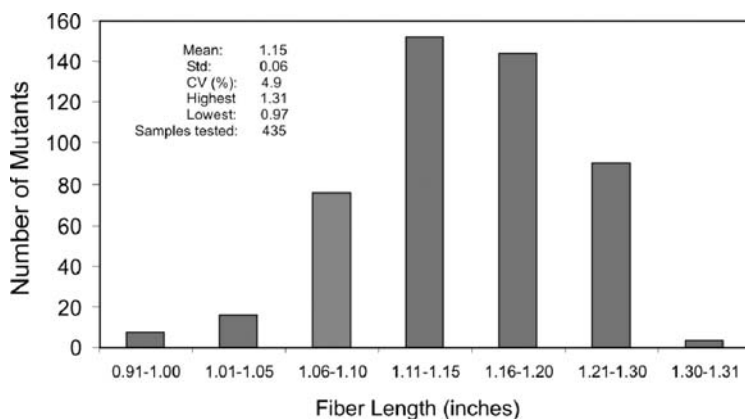


Fig. 1 Distribution of fiber length in 435 M₆ lines of FM 958 grown at Lubbock, Tx in 2006

lines with short fiber length (Ranging from 0.97 to 1.05 inches) and 18 M₆ lines with long fiber (ranging from 1.25 to 1.31 inches) were selected for progeny testing in 2007. It was interesting that mean of the 435 M₆ lines was almost identical to both the commercial and experimental fiber length expected of FM 958 in the 2006 production year at Lubbock, TX but the variation far exceeded the level expected of 435 randomly selected plants of FM 958. Genetic analyses of these mutants selected for both short and long fiber length should help us identify and eventually map alleles which impact fiber length in upland cotton.

4 Reverse Genetics via TILLING

The explosion of DNA sequence information has created new opportunities to investigate the function of plant genes. The approach by which genes are first identified by DNA sequence analysis and the biological function is subsequently derived through targeted gene manipulation is known as reverse genetics. The tools available for reverse genetics in plants include the use of transposons or T-DNA for gene tagging and the expression of antisense or hairpin RNAs that induce epigenetic silencing of homologous genes. While powerful, these methods have significant limitations. For example, they typically require gene transfer technology and, thus, have limited applicability beyond selected plant species. Also, although the insertion of T-DNAs and transposons into plant genomes is not site specific, it is also not random. There exist clear insertional “hotspots”, sites where insertions occur frequently, and “coldspots”, sites where insertions are rare (Pan et al., 2005). Thus, saturation mutagenesis using these insertional elements is difficult, as demonstrated by the fact that, in spite of mapping approximately 360,000 T-DNA insertions in the *Arabidopsis* genome, 10% of the genes remain untagged (SIGNAL T-DNA Express). RNA suppression

strategies also have limitations in that they require extensive development of gene constructs and plant transformation. In addition, variability in gene suppression can lead to difficulties in phenotypic analysis.

A complementary method for reverse genetics that depended on identification of specific mutants in a chemically mutagenized plant population was first described by McCallum et al. (2000a; 2003b). This method, known as **Targeting Induced Local Lesion IN Genomes (TILLING)**, is an effective and widely applicable reverse genetics approach. TILLING combines chemical mutagenesis with high-throughput screens for sequence mismatches in pooled PCR products. Chemical mutagens such as EMS and MNU are alkylating agents that affect specific nucleotide resulting in an essentially random distribution of nucleotide substitutions. Thus, even relatively small mutant populations can produce an allelic series that can include both missense and nonsense mutant alleles. TILLING is a technique designed to identify individual plants that have a mutation in a gene of interest whose sequence is known. TILLING has two significant advantages over existing plant gene knock-out tools: first, it is applicable to any plant since it does not require transgenic or cell culture manipulations; second, it produces an allelic series of mutation, including hypomorphic alleles that are useful for genetic analysis. TILLING has been used successfully to identify novel point mutation in targeted genes in M₂ EMS mutagenized *Arabidopsis* population (McCallum et al. 2000). In addition to screening mutagenized populations for induced mutations, TILLING can also be used to screen germplasm resources for natural DNA polymorphism. This process is known as “*Ecotilling*” (Comai et al., 2004).

The original TILLING technique described by McCallum et al. (2000) required the use of denaturing high performance liquid chromatography (dHPLC) to identify mismatched DNA fragments. In this approach, DNA samples were purified from individual plants of an EMS-mutagenized population. DNA samples from up to 8 individual plants were combined and the pooled DNA amplified by polymerase chain reaction (PCR) using primers that are specifically designed to amplify a 500 to 1000 bp region of the gene of interest. After the final amplification cycle, the amplified DNA sample was denatured at 95°C then slowly cooled to 25°C to allow for renaturation. If DNA polymorphisms existed between the 8 pooled DNA samples, the renaturation products would include heteroduplexes that contained mismatched nucleotides, in addition to the perfectly matched homoduplexes. Due to their nucleotide mismatches, heteroduplexes had reduced stability and could be detected by their lower melting points in dHPLC. While effective, this method allowed for the analysis of only one pooled DNA sample at a time and, therefore, could not be considered high-throughput.

A simpler TILLING method was subsequently developed that depended on the single strand-specific nuclease CEL I to detect mismatched DNA base pairs (Oleykowski et al., 1998). CEL I, a single strand-specific nuclease extracted from celery stalks, digests mismatched base-pairs in heteroduplexed DNA fragments with high efficiency and specificity and can, be used to detect

mutations (Kulinski et al., 2000; Till et al., 2004). This enzyme had become the nuclease of choice for TILLING. In this technique, the amplified products of the DNA pools are digested with CEL I and the products resolved by gel electrophoresis, the truncated DNA fragments that result from CEL I digestion of heteroduplexes that contain base pair mismatches being evident as novel bands within the electrophoretic pattern. Resolution of these samples using a LiCOR automated DNA sequencer system allows for simultaneous analysis of dozens of samples, making the process high-throughput (Colbert et al., 2001; Till et al., 2003a; 2003b)

Publically funded TILLING projects for several crop species have been established that offer reverse genetics services for researchers within the community. These include the Maize TILLING Project at Purdue University (<http://genome.purdue.edu/maizetilling/index.htm>), the rice and tomato TILLING projects at the University of California at Davis (http://tilling.ucdavis.edu/index.php/Main_Page), a soybean TILLING project (<http://www.soybeantilling.org/index.jsp>), and a Lotus TILLING project at the Sainsbury Laboratory and John Innes Centre (<http://www.lotusjaponicus.org/tilling/pages/homepage.htm>). Although a public Cotton TILLING Project has not yet been established, preliminary work to adapt the TILLING procedure for use in diploid cotton was carried out in our laboratories. DNA was extracted from approximately 1000 EMS-mutagenized M² *Gossypium arboreum* plants using a DNeasy plant DNA extraction kit (Qiagen) and equal amounts of DNA were pooled in groups of 8. Specific PCR primers for the brassinosteroid receptor gene *GhBR11* (Friedrichsen et al., 2000; Wang et al., 2001) were used to amplify an approximately 600 bp fragment of coding sequences selected for sequence characteristics where GC to AT substitution were likely to result in stop codons. *GhBR11* was selected because brassinosteroid has been shown to have a significant impact on cotton fiber development (Sun et al., 2003). Forward and reverse primers were labeled with LiCOR IR dye 700 and 800, respectively. Amplification of target DNA sequences in pooled DNA sample was carried out using high fidelity *Taq* polymerase and the amplification reactions were heated to 95°C then slowly cooled to 25°C to allow renaturation. Renatured samples were digested with CEL I and the digested samples loaded onto polyacrylamide slab gels on a LiCOR automated DNA sequencer. Fluorescence data was captured and analyzed visually.

Pooled reactions that included amplified DNA samples from plants with base substitutions in the *GhBR11* gene were identified by the presence of rapidly migrating bands that corresponded between both strands. Once these pools were identified, DNA samples from each individual plant within the pool were analyzed with DNA from a known wild-type plant. Amplified DNA fragments from individual plants that are found to contain point mutations within the targeted sequence were sequenced to identify the specific lesions. Two distinct mismatches were found in this analysis and both were silent G to A transitions at the third position of a codon. While G to A transitions are consistent with EMS mutagenesis, the high rate at which these mismatches within the selected

region of the *GhBR11* gene occurred (1 in approximately 250 plants) suggest that one or both could represent natural DNA polymorphisms within the *G. arboreum* population. Although we did not identify loss of function mutations in this limited screen, these unpublished results clearly show that the high-throughput TILLING procedures are readily applicable to cotton.

5 Insertional Mutagenesis

Insertional mutagenesis has proven to be an efficient strategy for direct isolation of genes by identifying their mutant phenotypes. The power of this method is due to the fact that the insertional mutagen has dual functions. It both alters gene function and provides sequence information for a direct recovery of the flanking DNA using PCR-based methods. Thus, genes affected by the mutagenic process can be quickly identified and characterized, obviating the need for genetic mapping and positional cloning. Two basic types of insertional mutagens have been used successfully in plants, transposable elements and T-DNA of *Agrobacterium*. T-DNA tagging has been used to identify and clone many genes of the model plant *Arabidopsis thaliana* (Bouche and Bouchez 2001) and rice (Jeong et al., 2002). However, T-DNA insertions tend to be single copy, which requires a large population to be screened, making this strategy appropriate only for plants such as *Arabidopsis* with high through-put transformation methods. Transposon-based systems, on the other hand, can produce a large number of insertions from a few initial transformation events because these elements are able to transpose. Therefore, the likelihood of generating new insertion events is high (Dilkes and Feldmann 1998).

Gene tagging by endogenous transposable elements (TEs) has been used for decades to isolate genes from maize and it is estimated that in few years, on-going genomics studies will define all maize genes using transposon insertion sites (Walbot, 2000 and 1992). The same maize elements proved to be functional in heterologous plant systems (Osborne and Baker 1995). The first report showing an autonomous transposition of maize *Activator* element in tobacco (Baker et al. 1986) opened the door for an extensive use of maize transposable elements in other plant species. So, in less than a decade, TEs from maize were shown to be active in several plant species including *Arabidopsis* (Aarts et al. 1993; Bancroft and Dean 1993), tobacco (Masson and Fedoroff, 1989), tomato (Meissner et al. 2000; Keddie et al. 1998; Takken et al. 1998; van der Biezen et al. 1996), petunia (Robbins et al. 1994; Chuck et al. 1993) flax (Anderson et al. 1997), carrot (Van Sluys et al. 1987), potato (Knapp et al. 1988; Frey et al. 1989), rice (Kohli et al. 2001; Murai et al. 1991), barley (Koprek et al. 2001; Scholz et al. 2001) soybean (Zhou and Atherly 1990) and wheat (Laufs et al. 1990).

Although, there are at least eight independent families of TEs in maize (Peterson, 1987) two systems have been widely used in heterologous hosts. These are the *Activator/Dissociation* (*Ac/Ds*) in the *Ac* super-family and the

Enhancer/Inhibitor (En/I) belonging to the CACTA super-family. The *En/I* is also known as *Suppressor/mutator (Spm/dSpm)* (Kunze et al. 1997). Both systems include autonomous elements (or master elements), *Ac* and *En*, that encode a transposase (TPase) required for mobilization of the non-autonomous (or receptor) elements, *Ds* and *I*. The receptor elements are usually deletion derivatives of the autonomous elements that cannot transpose by themselves because they have lost the TPase gene but retain the terminal sequences required for transposition.

The well-characterized *Ac/Ds* transposon system has been widely used for gene tagging in several heterologous dicots and monocots plants (Bhatt et al. 1996; Kunze et al. 1997). The *cis*-determinant for transposition is 11-bp terminal inverted repeat sequences (TIRs) plus a 240-bp located at the sub-terminal region at both ends (Coupland et al. 1989; Kunze and Starlinger. 1989, Pohlman et al. 1984). The *Ac* element codes for a 3.5-kb TPase that is essential for transposition (Heinlein et al. 1994; Kunze et al. 1993). The interaction of TPase and TIRs in *trans* result in the excision of the element in a cut-and-paste manner (Kunze et al. 1997). The excision is usually not precise, which leaves behind a target site duplication (footprint) or minor sequence alterations such as small deletions or other changes. Although *Ac* is active in many different plant species, its transposition rate varies considerably (Kunze, 1996). This variation in transposition was attributed to the variation of TPase activity in different host plants. Unlike in maize where high concentration of TPase negatively affects the transposition resulting in an "inverse dose effect" (McClintock, 1951; Kunze et al. 1997) a direct positive relationship between TPase and transposition frequency was reported in *Arabidopsis* (Dean et al. 1992) and tobacco (Keller et al. 1993a). The choice of the promoter driving the transposase source was also shown to be important for efficient transposition. For example, in tobacco and tomato the native TPase promoter resulted in suitable transposition frequency, whereas in *Arabidopsis* the strong CaMV 35S promoter was required (Grevelding et al. 1992; Long et al. 1993). The CaMV 35S promoter induced excision early in development that produced large somatic sectors (Long et al. 1993). Except in tomato where *Ac* and *Ds* transpose to linked and unlinked sites with equal frequency, in all tested plant the *Ac* and *Ds* tend to favor genetically linked sites with insertions into these sites occurring up to 70% of the time. (Bancroft and Dean 1993; Keller et al. 1993b).

To improve transposon tagging using *Ac/Ds* elements, a two components system was developed and proved to be useful, at least in *Arabidopsis*. This system consists of a stable TPase source that cannot transpose and a *Ds* element that will transpose only if transposase is provided in *trans*. The *Ac* and *Ds* elements are separately introduced in different host plants and mutagenesis is then initiated by crossing the two lines. The transposase source can be under the control of strong promoter such as CaMV 35S, inducible promoter or native promoter. Such manipulations are important to determine the best conditions for efficient transposition. The *Ds* element can also be engineered to carry reporter genes and/or selectable markers to monitor the excision/integration

events and expression of plant genes adjacent to the transposon insertion site (Springer et al. 1995). Since transposed *Ds* elements tend to insert at locations genetically linked to their starting position, strategies have been developed to select for plant with unlinked insertion. This involved selecting against both the T-DNA harboring the transposase source and the T-DNA from which *Ds* has been excised at the same time selecting for transposed *Ds* element (Springer et al. 1995, Sundaresan et al. 1995). This counter selection was possible by using various dominant-negative selectable markers such as *tms2* from *Agrobacterium* (Springer et al. 1995), *codA* from *E. coli* (Koprek et al. 1999; Perera et al. 1993) and *CYP105A1* from *Streptomyces griseolus* (Koprek et al. 1999; O'Keefe et al. 1994).

In the *En/I* system, two components are necessary for transposition, a 13-bp terminal inverted repeat sequences (TIRs), characteristic of CACTA superfamily and *TnpA/TnpD* gene products encoded by the autonomous *En* element (Preira and Aarts 1998; Schiefelbein et al. 1988). Both in maize and other plants, the *En/I* tends to create three-bp duplications at the insertion sites leaving characteristic footprints upon imprecise excision (Preira and Aarts 1998). Cardon and coworkers (1993) reported that in *Arabidopsis*, germinal excisions and reinsertions proceed in a similar manner as in maize. However, when CaMV 35S promoter was used to drive the *En* genes instead of the native promoter, excision events occurred early during development (Masson and Fedoroff, 1989). Aarts and coworkers (1995a) developed a modified *En/I* system for transposon tagging in *Arabidopsis*. This system consists of a non-autonomous, stable *En* element under the control of CaMV 35S promoter and an unmarked *I* element inserted into the open reading frame of an *NPTII* (neomycin phosphotransferase) gene to monitor excision phenotypically. Continuous transposition throughout plant development was detected over at least seven consecutive generations resulting in the generation of several independent lines with unique transposition events. The frequency of independent transposition ranges between 7-29% in different progeny. Subsequently, several genes were tagged and isolated using the *En-I* system. This includes *MS2* male sterility (Aarts et al. 1993), *CER1* wax biosynthesis (Aarts et al. 1995b), *F3H* involved in flavonoid biosynthesis (Wisman et al. 1998), *FDH* involved in fatty acid elongation (Yephremov et al. 1999) and *LCR* a fatty acid ω -hydroxylase involved in cutin biosynthesis (Wellesen et al. 2001).

More recently, an activation tagging system using the *En-I* was used in *Arabidopsis* (Marsch-Martinez et al. 2002). In this case, the *I* element harbors a tetramer of the CaMV 35S enhancer plus a *Bar* gene for positive selection of the inserts. A negative selectable marker, *SU1*, was introduced adjacent to the *I* element to allow for selection against linked insertions. The advantage of this system is the ability to generate gain-of-function dominant mutations. The dominant mutant frequency was estimated to be 1% and several novel gain-of-function mutations were identified (Marsch-Martinez et al. 2002). Because the of high frequency of mutant and the small number of primary transformants required, the authors suggest that this

system will provide opportunities for activation tagging in plants with low transformation efficiency (Marsch-Martinez et al. 2002).

Unlike AC/DS and EN/I transposons, LTR-retrotransposons transpose via an RNA intermediate. These RNAs are transcribed from the DNA element, reverse transcribed in the cytoplasm within virus-like particles to generate double-stranded DNA intermediates that are imported into the nucleus and integrated in the host genome. Retrotransposons are found in virtually all eukaryotic genomes, including plants (Voytas et al., 1992; Suoniemi et al., 1996) and often constitutes large portions of the genomic mass. For example, the *BARE-1* element accounts for 7% of the barley genome and at least 50% of the maize genome consists of retrotransposons of various families (San Miguel et al., 1996). Transcription of several plant retrotransposons has been shown to be induced by stress and, in particular, tissue culture and transposition of the elements *Tnt1* and *Tto1* is induced by tissue culture in their native tobacco host (Hirochika, 1993). Although insertion sites of plant retrotransposons are not well characterized, these elements tend to accumulate in heterochromatic regions of many species. For example, intergenic clusters of retrotransposons are found in the maize genome (San Miguel et al., 1996); however mutations arising from retrotransposon insertions have been isolated in maize (Johns et al., 1985; Varagona et al., 1992), indicating that new insertion events can occur in genes and the *Tos17* transpositions in rice were shown to occur in transcribed regions (Hirochika et al., 1996).

Tnt1 is one of the few plant retrotransposons that have been shown to undergo autonomous transposition in other plant species (Lucas et al. 1995, Feuerback et al., 1995). Transposition of *Tnt1* in *Arabidopsis* results in insertions within genes (Courtial et al., 2001) and the *Tnt1* system has been successfully used to tag genes in lettuce (Mazier et al., 2007) and the model legume *Medicago truncatula* (Tadege et al., 2008). In More than 7,600 independent lines were developed in *M. truncatula*, representing approximately 190,000 insertion events. On average, *Tnt1* inserted at 25 different locations per genome and the insertions were stable during subsequent generation. Analysis of sequences flanking *Tnt1* insertions revealed that *Tnt1* appears to insert preferentially into gene-rich regions with no apparent target site specificity. Screening of 3237 R₁ lines resulted in identification of visible mutant phenotypes in approximately 30% of the regenerated line and all of the examined were found to be tagged. The authors estimated that approximately 16,000 lines would be sufficient to tag 90% of the genes in this species. In contrast, more than 500,000 lines would be required to achieve the same level of saturation using T-DNA tagging. Thus, *Tnt1* appears to be an efficient tool for insertional mutagenesis in plant species with large genomes.

Due to the fairly recent evolution of tetraploid cotton (Paterson 2001; Wendel, 1989), and the fact that cultivated cottons have been heavily selected for agronomic traits, the genetic diversity of cotton is quite low (Small et al., 1999). This lack of genetic diversity diminishes our ability to understand the basic mechanisms that regulate cotton development and limits the application

of molecular genetic tools for effective improvement of cotton germplasm. Development of insertional mutagenesis resources in cotton could be used to increase allelic diversity and provide expanded genetic resources for genetic research and improvement of cotton germplasm. While procedures for the genetic transformation of cotton using *Agrobacterium*-based methods and regeneration of transgenic plants via somatic embryogenesis are well established, these methods do not provide the necessary through-put to efficiently develop T-DNA tagged populations. Therefore, research in our laboratory has focused on transposon-based strategies for activation tagging and gene/enhancer trap methods that will allow researchers to identify interesting new genes while avoiding problems with gene redundancy that plague efforts to identify recessive alleles in polyploid species. Mutagenized populations will be established that can be screened for novel phenotypes affecting virtually all plant process including growth and development, reproductive characteristics, or resistance to stress or pests. While these processes are not unique to cotton, it is possible that new insight can be gained by evaluating these phenotypes in cotton. On the other hand, cotton is the only plant that produces trichomes that can be spun into yarn for textiles. Therefore, cotton is the de facto model for this developmental process. For this reason, screening of these mutagenized populations is likely to focus on genes that affect fiber development processes. We anticipate that this approach can lead to the identification of large numbers of novel genes that will provide important insights specifically into cotton fiber differentiation and into cell elongation, cell wall synthesis, and cellular maturation in general.

6 Conclusions

Cotton is amenable to currently available mutagenesis procedures. Use of chemicals and radiation have been used to successfully increase genetic variability in cotton and the feasibility of using TILLING to identify DNA-based polymorphisms has been demonstrated. Initial steps to develop insertional mutagenesis procedures are now underway. We anticipate that these approaches will provide important genetic tools for the future exploration of gene function in cotton.

Acknowledgments The authors thank Haggag Abdel-Mageed and Bay Nguyen for their work on transposon mutagenesis and TILLING, respectively. The authors acknowledge the generous support from Cotton Inc.

References

- Aarts MG, Dirkse WG, Stiekema WJ, Pereira A. 1993. Transposon tagging of a male sterility gene in *Arabidopsis*. *Nature*. 363: 715–717.
- Aarts MG, Corzaan P, Stiekema WJ, Pereira A. 1995a. A two-element Enhancer/Inhibitor transposon system in *Arabidopsis thaliana*. *Mol Gen Genet*. 247: 555–564.

- Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A. 1995b. Molecular characterization of the *CER1* gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell*. 7: 2115–2127.
- Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG. 1997. Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell*. 9: 641–651.
- Akhtar KP, Aslam M, Haq M, Jamil FF, Khan AI, Elahi MT (2005) Resistance to cotton leaf curl virus (CLCuV) in a mutant cotton line. *J Cotton Sci* 9:175–181.
- Aslam M, Elahi MT, Iqbal N (2003) Development of improved germplasm of cotton through radiation and DNA-mediated embryo transformation technique - evaluation and confirmation of novel genotypes. In: Improvement of New and Traditional Industrial Crops by Induced Mutations and Related Biotechnology. International Atomic Energy Agency pp. 69–80.
- Aslam M, Iqbal N, Bandesha AA, Haq MA (2004) Inductions of mutations through crosses with gamma irradiated pollen in cotton. *Inter J Agric Biol* 6:894–897.
- Auld DL, Heikkinen M.K., Erickson DA, Sernyk, JL, Romero JE. (1992) Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. *Crop Sci* 32:357–362.
- Auld, D.L., M.D. Ethridge, J.K. Dever and P.D. Dotray. (1998) Chemical mutagenesis as a tool in cotton improvement. P. 550-551. In: P. Duggar and D.A. Richter (eds.) *Proc. Beltwide Cotton Conf.* San Diego, CA. 5-9 Jan. Natl. Cotton Council, Memphis, TN.
- Auld D.L., Bechere E, Ethridge MD, Becker WD, Hequet E, Cantrell, RG (2000) Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. *Crop Sci*. 40:1835–1836.
- Baker B, Fedoroff N., Loerz, H., Schell J. 1986. Transposition of the maize controlling element activator in tobacco. *Proc. Natl. Acad. Sci. USA*. 83: 4844–4848.
- Bancroft I, Dean C. 1993. Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana*. *Genetics*. 134: 1221–1229.
- Bhatt AM, Page T, Lawson EJ, Lister C, Dean C. 1996. Use of *Ac* as an insertional mutagen in *Arabidopsis*. *Plant J*. 9: 935–945.
- Bouche N, Bouchez D. 2001. *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant. Biol.* 4: 111–117.
- Bowman, D.T., O.L. May, and D.S. Calhoun. (1996) Genetic Base of Upland Cotton Cultivars Released between 1970 and 1990. *Crop Sci*. 36:577–581.
- Cardon GH, Frey M, Saedler H, Gierl A. 1993. Mobility of the maize transposable element *En/Spm* in *Arabidopsis thaliana*. *Plant J*. 3: 773–784.
- Chuck G, Robbins T, Nijjar C, Ralston E, Courtney-Gutterson N, Dooner HK. 1993. Tagging and cloning of a petunia flower color gene with the maize transposable element *Activator*. *Plant Cell* 5: 371–378.
- Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S. (2001) High-throughput screening for induced point mutations. *Plant Physiol*. 126: 480–484.
- Comai L, Young K, Reynolds SH, Codomo C, Enns L, Johnson J, Burtner C, Henikoff JG, Greene EA, Till BJ, Henikoff S. (2004) Efficient discovery of nucleotide polymorphisms in populations by ecotilling. *Plant Journal*. 37:778–786.
- Comai L, Henikoff S (2006) TILLING: practical single-nucleotide mutation discovery. *Plant J* 45:684–694.
- Comis D (2005) TILLING genes to improve soybeans. *Agric Res July*:4-5.
- Cornelius TJ, Peter SD, Narayanan SS, Kamalanathan S (1970) Useful mutants in Cambodia cotton (*G. hirsutum*) by ionising radiation of seeds. *Madras Agricultural J* 57:594–597.
- Coupland G, Plum C, Chatterjee S, Post A, Starlinger P. 1989. Sequences near the termini are required for transposition of the maize transposon *Ac* in transgenic tobacco plants. *Proc. Natl. Acad. Sci. U S A*. 86: 9385–9388.

- Courtial B, Feuerbach F, Eberhard S, Rohmer L, Chiapello H, Camilleri C, Lucas H. 2001. *Tnt1* transposition events are induced by in vitro transformation of *Arabidopsis thaliana*, and transposed copies integrate into genes. *Mol Genet Genomics*. 265: 32–42.
- Dean C, Sjodin C, Page T, Jones JDG, Lister C. 1992. Behavior of the maize transposable element *Ac* in *Arabidopsis thaliana*. *The Plant Journal*. 2: 69–81.
- Dilkes BP, Feldmann KA. 1998. Cloning genes from T-DNA tagged mutants. *Methods Mol Biol*. 82: 339–351.
- Feuerbach F, Drouaud J, Lucas H. 1997. Retrovirus-like end processing of the tobacco *Tnt1* retrotransposon linear intermediates of replication. *J Virol*. 71: 4005–4015.
- Frey M, Tavantzis, SM, Saedler, H. 1989. The maize En-1/Spm element transposes in potato. *Molecular Genl Genet*. 21 : 172–177.
- Friedrichsen DM, Joazeiro CA, Li J, Hunter T, Chory J. (2000) Brassinosteroid-insensitive –1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant physiol*. 123: 1247–1256
- Fotiadis NA, Miller PA (1973) Effects of recurrent seed irradiation on genetic variability and recombination in cotton (*Gossypium hirsutum* L.). *Crop Sci* 13:40–44.
- Gaibullaev I K, Avazkhodzhaev M K, Egamberdiev AE (1975) Reaction of chemomutants of cotton to infection of the pathogen of verticillium wilt. *Soviet Genetics* 11:34–36.
- Gao W, Chen ZJ, Yu JZ, Raska D, Kohel RJ, Womack JE, Stelly DM (2004) Wide-cross whole-genome radiation hybrid mapping of cotton (*Gossypium hirsutum* L.). *Genetics* 167:1317–1329.
- Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. *Current Opinion in Plant Biology* 8:211–215.
- Greveling C, Becker D, Kunze R, von Menges A, Fantes V, Schell J, Masterson R. 1992. High rates of *Ac/Ds* germinal transposition in *Arabidopsis* suitable for gene isolation by insertional mutagenesis. *Proc. Natl. Acad. Sci. USA*. 89: 6085–6089.
- Gulin, VV, Kal'chenko VA, Turavekov S, Musaev DA (1985) Polymorphism and mutagenesis of nonspecific esterases in cotton (*Gossypium hirsutum* L.). *Doklady: Biol Sci* 285:700–703.
- Heinlein M, Brattig T, Kunze R. 1994. In vivo aggregation of maize Activator (*Ac*) transposase in nuclei of maize endosperm and *Petunia* protoplasts. *Plant J*. 5: 705–714.
- Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. *Plant Phys*. 135:630–636.
- Herring AD, Auld DL, Ethridge MD, Hequet EF, Bechere E, Green CJ, Cantrell RG (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. *Euphytica* 136:333–339.
- Hirochika H. 1993. Activation of tobacco retrotransposons during tissue culture. *EMBO J*. 12: 2521–2528.
- Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S. 1996. Autonomous transposition of the tobacco retrotransposon *Tio1* in rice. *Plant Cell* 8: 725–734.
- Hohmann U, Jacobs G, Jung C (2005) An EMS mutagenesis protocol for sugar beet and isolation of non-bolting mutants. *Plant Breeding* 124:317–321.
- Horlacher WR, Killough DT (1933) Progressive mutations induced in *Gossypium hirsutum* by radiations. *Amer Naturalist* 67:532–538.
- Hussein HAS, Al-enani A, El-Moghazi M (1982) Histological and morphological characteristics of glandless cotton mutant induced with sodium azide. *Egyptian Journal of Genetics and Cytology* 11:167–173.
- Hutchinson, J.B., Silow, R.A., and Stephens, S.G. (1947) In "The evolution of *Gossypium* and the differentiation of the cultivated cottons". Oxford University Press, London.
- James DW Jr., Dooner H K (1990) Isolation of EMS-induced mutants in *Arabidopsis* altered in seed fatty acid composition. *Theor Appl Genet* 80:241–245.
- Jeong D-H, An S, Kang H-G, Moon S, Han J-J, Park S, Lee H-S, An K, and An G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol*. 130: 1636–1644.

- Johns MA, Mottinger J, Freeling M. 1985. A low copy number, copia-like transposon in maize. *EMBO J.* 4: 1093–1102.
- Kandhro MM, Laghari S, Sial MA, Nizamani GS (2002) Performance of early maturing strains of cotton (*Gossypium hirsutum* L.) developed through induced mutation and hybridization. *Asian J Plant Sci* 5:581–582.
- Katterman FRH (1973) 5-Methyl cytosine content in the DNA of colchicine and spontaneously induced polyhaploids of *Gossypium*. *Phytochemistry* 12:1887–1889.
- Keddie JS, Carroll BJ, Thomas CM, Reyes ME, Klimyuk V, Holtan H, Gruissem W, Jones JD. 1998. Transposon tagging of the Defective embryo and meristems gene of tomato. *Plant Cell* 10: 877–888.
- Keller J, Jones JD, Harper E, Lim E, Carland F, Ralston EJ, Dooner HK. 1993a. Effects of gene dosage and sequence modification on the frequency and timing of transposition of the maize element Activator (*Ac*) in tobacco. *Plant Mol. Biol.* 21: 157–170.
- Keller J, Lim E, Dooner HK. 1993b. Preferential transposition of *Ac* to linked sites in *Arabidopsis*. *Theoretical and Applied Genetics* 86: 585–588.
- Knapp S, Coupland G, Uhrig H, Starlinger P, Salamini F. 1988. Transposition of the maize transposable element *Ac* in *Solanum tuberosum*. *Mol. Gen. Genet.* 213: 285–290.
- Kohli A, Xiong J, Greco R, Christou P, Pereira A. 2001. Tagged Transcriptome Display (TTD) in *Indica* rice using *Ac* transposition. *Mol Genet Genomics*. 266: 1–11.
- Koprek T, McElroy D, Louwerse J, Williams-Carrier R, Lemaux PG. 1999. Negative selection systems for transgenic barley (*Hordeum vulgare* L.): comparison of bacterial *codA*- and cytochrome P450 gene-mediated selection. *Plant J.* 19: 719–726.
- Koprek T, Rangel S, McElroy D, Louwerse JD, Williams-Carrier RE, Lemaux PG. 2001. Transposon-mediated single-copy gene delivery leads to increased transgene expression stability in barley. *Plant Physiol.* 125: 1354–1362.
- Kulinski, J, Besack, D, Oleykowski, CA, Godwin, AK and Yeung, AT. 2000. CEL I enzymatic mutation detection assay. *Biotechniques*. 29: 44–46.
- Kunze R, Starlinger P 1989. The putative transposase of transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*. *EMBO J.* 8: 3177–3185.
- Kunze R, Behrens U, Courage-Franzkowiak U, Feldmar S, Kuhn S, Lutticke R. 1993. Dominant transposition-deficient mutants of maize Activator (*Ac*) transposase. *Proc. Natl. Acad. Sci. USA.* 90: 7094–7098.
- Kunze R. 1996. The activator (*Ac*) element of *Zea mays* L. In “Transposable Elements” (H. Saedler and A. Gierl, eds), pp. 161–194. Springer, Heidelberg.
- Kunze, R, Saedler H., Lonnig WE. 1997. Plant transposable elements. *Botanical Research* 27, 331–470.
- Larik AS, Hafiz HMI, Al-Saheal YA (1983) Azide mutagenesis in cotton (*Gossypium hirsutum*). *Science and Environment* 5:33–42.
- Laufs J, Wirtz U, Kammann M, Matzeit V, Schaefer S, Schell J, Czernilofsky AP, Baker B, Gronenborn B. 1990. Wheat dwarf virus *Ac/Ds* vectors: expression and excision of transposable elements introduced into various cereals by a viral replicon. *Proc. Natl. Acad. Sci. USA.* 87: 7752–7756.
- Long D, Swinburne J, Martin M, Wilson K, Sundberg E, Lee K, Coupland G. 1993. Analysis of the frequency of inheritance of transposed *Ds* elements in *Arabidopsis* after activation by a *CaMV* 35S promoter fusion to the *Ac* transposase gene. *Mol. Gen. Genet.* 241: 627–636.
- Lucas H, Feuerbach F, Kunert K, Grandbastien MA, Caboche M. 1995. RNA-mediated transposition of the tobacco retrotransposon *Tnt1* in *Arabidopsis thaliana*. *EMBO J* 14: 2364–2373
- Luckett DJ (1989) Colchicine mutagenesis is associated with substantial heritable variation in cotton. *Euphytica* 42:177–182.
- Marsch-Martinez N, Greco R, Van Arkel G, Herrera-Estrella L, Pereira A. 2002. Activation tagging using the *en-1* maize transposon system in *Arabidopsis*. *Plant Physiol.* 129: 1544–1556.

- Masson P, Fedoroff NV. 1989. Mobility of the maize suppressor-mutator element in transgenic tobacco cells. *Proc. Natl. Acad. Sci. USA*. 86: 2219–2223.
- May, O.L., D.T. Bowman, and D.S. Calhoun. (1995) Genetic diversity of U.S. Upland Cotton Cultivars Released between 198- and 1990. *Crop Sci.* 35:1570–1574.
- Mazier M, Botton E, Flamain F, Bouchet J-P, Courtial B, Chupeau M-C, Chupeau Y, Maisonneuve B, Lucas H. 1886 Successful gene tagging in lettuce using the *Tnt1* retro-transposon from tobacco. *Plant Physiol.* 144: 18–31.
- McCallum CM, Comai L, Greene EA, Henikoff S. (2000) Targeted screening for induced mutations. *Nature Biotechnology*. 18: 455–457.
- McCallum CM, Comai L, Green EA, Henikoff S. (2000) Targeting induced local lesion in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123: 439–442.
- McClintock B. 1951. Chromosome organization and genetic expression. Cold Spring Harbor Symposia on Quantitative Biol. 16: 13–47.
- Mehetre SS, Thombre MV (1982) Fibre properties of x-ray induced glandless mutants in American cotton. *J Maharashtra Agricultural Universities* 8:189–190.
- Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy AA. 2000. Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato. *Plant J.* 22: 265–274.
- Murai N, Li ZJ, Kawagoe Y, Hayashimoto A. 1991. Transposition of the maize activator element in transgenic rice plants. *Nucleic Acids Res.* 19: 617–622.
- National Cotton Council. (2005) Economic Impact of US Cotton. www.cotton.org/econ/cropinfo/supply-demand.cfm
- Nazirov NN, Satipov G (1979) Reaction of radiation induced mutants to different water regimes. *Soviet Agricultural Sciences* 7:12–15.
- Nazirov NN, Tashmatov NT, Bakhabov A, Nabiev AG (1981) Respiration rate and rate of ³²P incorporation into organophosphorus compounds in radiation-induced cotton mutants and their initial forms on saline soil. *Soviet Agricultural Sciences* 5:27–29.
- Ngemtov M, Kovalenko VI, Shumnyi VK, Asrorov KA (1975) Induction of cytoplasmic male sterility in cotton by the method of radiation mutagenesis. *Soviet Genetics* 11:1593–1595.
- Oleykowski CA, Mullins CRB, Godwin AK, Yeung AT. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*. 26: 4597–4602.
- O'Keefe DP, Tepperman JM, Dean C, Leto KJ, Erbes DL, Odell JT. 1994. Plant expression of a bacterial cytochrome P450 that catalyzes activation of a sulfonylurea pro-herbicide. *Plant Physiol.* 105: 473–482.
- Osborne BI, Baker B. 1995. Movers and shakers: maize transposons as tools for analyzing other plant genomes. *Curr. Opin. Cell Biol.* 7: 406–413..
- Pan X, Li Y, Stein, L. (2005) Site preferences of insertional mutagenesis agents in Arabidopsis. *Plant Physiol.* 137: 168–175.
- Pereira A, Aarts MG. 1998. Transposon tagging with the *En-I* system. *Methods Mol. Biol.* 82: 329–338.
- Perera RJ, Linard CG, Signer ER. 1993. Cytosine deaminase as a negative selective marker for Arabidopsis. *Plant Mol. Biol.* 23: 793–799.
- Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Phys.* 131:866–871.
- Peterson, P.A. 1987. Mobile elements in plants. *Critical Reviews in plant Sciences* 6: 105–208.
- Pohlman RF, Fedoroff NV, Messing J. 1984. Correction: nucleotide sequence of *Ac*. *Cell*. 39: 417.
- Raut RN, Jain HK, Panwar RS (1971) Radiation-induced photoinsensitive mutants in cotton. *Current Science* 40:383–384.
- Reddy VRK, Sundaravivelu K (1999) Improvement of cotton (*Gossypium hirsutum* L.) - Induced mutagenesis. In: *Proc Natl Symp Recent Trends in Plant Cytogenetics and Biotechnology*, 21-22 Feb, 1999. Dept of Botany, Andhra University, Visakhapatnam, India. Abst. No. 22.

- Robbins TP, Jenkin M, Courtney-Gutterson N. 1994. Enhanced frequency of transposition of the maize transposable element Activator following excision from T-DNA in *Petunia hybrida*. *Mol. Gen. Genet.* 244: 491–500.
- Salanki MS, Parameswarappa R (1968) Colchicine-induced mutant in cotton (*Gossypium hirsutum* L.) *Current Sci* 12:356–357.
- Sanamayan MF, Rakhmatullina RM (2003) Cytogenetic analysis of translocations in cotton. *Plant Breeding* 122:511–516.
- San Miguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.
- Savaskan C (2002) The effects of gamma radiation of the pollen size of *Gossypium hirsutum* L. *Turk J Biol* 26:477–480.
- Scholz S, Lorz H, Lutticke S. 2001. Transposition of the maize transposable element *Ac* in barley (*Hordeum vulgare* L.). *Mol Gen Genet.* 264: 653–661.
- Shamsuzzaman KM, Hamid MA, Azad MAK, Hussain M, Majid MA (2003) Varietal improvement of cotton (*Gossypium hirsutum*) through mutation breeding. In: Improvement of New and Traditional Industrial Crops by Induced Mutations and Related Biotechnology. International Atomic Energy Agency pp. 81–94.
- Shattuck V, Katterman FR (1982) Enhanced unscheduled DNA synthesis in the cotyledons of *Gossypium barbadense* L. by ethylmethanesulfonate (EMS). *Biochemical and Biophysical Research Communications* 109:1017–1025.
- Sheidai M, Azarani H, Hosseini Z (2002) Cytogenetic study of gamma irradiated lines of cotton (*Gossypium hirsutum* L.). *J Sci Islamic Rep of Iran* 13:311–322.
- Shi-Qi A, De-Qi Q, Xiu-Yun C (1991) Induction of parthenogenesis and chromosome behavior in plants of parthenogenetic origin in cotton (*Gossypium hirsutum*). *Genome* 34:255–260.
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotech* 23:75–80.
- Spasibionek S (2006) New mutants of winter rapeseed (*Brassica napus* L.) with changed fatty acid composition. *Plant Breeding* 125:259–267.
- Springer PS, McCombie WR, Sundaresan V, Martienssen RA. 1995. Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in Arabidopsis. *Science*. 268: 877–880.
- Sundaresan V, Springer P, Volpe T, Haward S, Jones JD, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* 9: 1797–1810.
- Sun Y, Veerabomma S, Abdel-Mageed HA, Fokar M, Asami T, Yoshida S, Allen RD. (2005) Brassinosteroid regulates fiber development on cultured cotton ovules. *Plant Cell Physiol.* 46: 1384–1391.
- Suoniemi A, Ananthawat-Jónsson K, Arna T, Schulman AH. 1996. The *BARE-1* retrotransposon is transcribed in barley from an LTR promoter active in transient assays. *Plant Mol Biol.* 31: 295–306.
- Takken FL, Schipper D, Nijkamp HJ, Hille J. 1998. Identification and Ds-tagged isolation of a new gene at the Cf-4 locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. *Plant J.* 14: 401–411.
- Tadege M, Wen Ji, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M, Ratet P, Mysore KS. 2008. Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal*, 54: 335–347.
- Till BJ, Colbert T, Tompa R, Enns LC, Codomo CA, Johnson JE, Reynolds SH, Henikoff JG, Greene EA, Stein MN, Comai L, Henikoff S. (2003a) High-throughput TILLING for functional genomics. *Plant Functional Genomics: Methods and Protocols* Edited by: Grotewald E. Clifton, NJ, Humana Press; pp. 205–220.

- Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S. (2003b) Large-scale discovery of induced point mutations with high throughput TILLING. *Genome Res*, 13: 524–530.
- Till BJ, Burtner C, Comai L, Henikoff S. 2004. Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*, 32:2632–2641.
- Tonnemaker, K.A., D.L. Auld, D.C. Thill, C.A. Mallory-Smith, and D.A. Erickson. (1992) Development of Sulfonylurea Resistant Rapeseed Using Chemical Mutagenesis. *Crop Sci*, 32:1387–1391.
- Van der Biezen EA, Brandwagt BF, van Leeuwen W, Nijkamp HJ, Hille J. 1996. Identification and isolation of the *FEEBLY* gene from tomato by transposon tagging. *Mol. Gen. Genet*, 251: 267–280.
- Van Sluys MA, Tempe J, Fedoroff N. 1987. Studies on the introduction and mobility of the maize Activator element in *Arabidopsis thaliana* and *Daucus carota*. *EMBO J*, 6: 3881–3889.
- Varagona MJ, Purugganan M, Wessler SR. 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell*, 4: 811–820.
- Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermeier SR. 1992. *cop*ia-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA*, 89: 7124–7128.
- Walbot V. 1992. Strategies for Mutagenesis and Gene Cloning Using transposon tagging and T-DNA insertional mutagenesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* Vol. 43: 49–82.
- Walbot V. Saturation mutagenesis using maize transposons. 2000. *Curr. Opin. Plant Biol*, 3: 103–107.
- Wang ZY, Seto H, Fujioka S, Yoshida S, Chory J. (2001) *BRI1* is a critical component of a plasma-membrane receptor for plant steroids. *Nature* 410: 380–3.
- Wellesen K, Durst F, Pinot F, Benveniste I, Nettesheim K, Wisman E, Steiner-Lange S, Saedler H, Yephremov A. 2001. Functional analysis of the *LACERATA* gene of *Arabidopsis* provides evidence for different roles of fatty acid omega hydroxylation in development. *Proc. Natl. Acad. Sci. USA*, 98: 9694–9699.
- Wendel, J.F. (1989) New World Tetraploid Cottons Contain Old-World Cytoplasm. *Proc Natl Acad Sci USA* 86: 4132–4136.
- Wisman E, Hartmann U, Sagasser M, Baumann E, Palme K, Hahlbrock K, Saedler H, Weisshaar B. 1998. Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc Natl Acad Sci U S A*, 95: 12432–12437.
- Wu J-L, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MRS, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ, Bruskiewich R, Wang G, Leach J, Khush G, Leung H (2005) Chemical- and irradiation induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol* 59:85–97.
- Xanthopoulos FP, Kechagia UE (2001) Improvement of two locally adapted cultivars in earliness by induced mutations. *Aust. J. Agric. Res.* 52:523–527.
- Xanthopoulos FP, Kechagia UE (2003) Improvement of two locally adapted cultivars by induced mutations. In: *Improvement of New and Traditional Industrial Crops by Induced Mutations and Related Biotechnology*. International Atomic Energy Agency pp. 61–68.
- Yakabova MM, Rubin AB, Khranova GA, Matorin DN (1975) Hill reaction and delayed fluorescence in mutants of *Gossypium hirsutum*. In: *Genetics Aspects of Photosynthesis*, Nasyrov et al., Eds. pp. 263–269.
- Yephremov A, Wisman E, Huijser P, Huijser C, Wellesen K, Saedler H. 1999. Characterization of the *FIDDLEHEAD* gene of *Arabidopsis* reveals a link between adhesion response and cell differentiation in the epidermis. *Plant Cell*, 11: 2187–2201.
- Zhou JH, Atherly AG. 1990. In situ detection of transposition of the maize controlling element (*Ac*) in transgenic soybean tissues. *Plant Cell Reports* 8: 542–545.

Gossypium Bioinformatics Resources

Alan R. Gingle

Abstract The primary goal of this chapter is to provide practical information for utilizing the array of *Gossypium* bioinformatics resources that are presently available. To establish the setting, the chapter begins with the description of a survey of *Gossypium* bioinformatics resources that was undertaken by the author in early 2007. Resources are categorized by life science area(s), available data types and available modes of data access. Navigating resources and searching for *Gossypium* data is then described through a broad collection of search examples that cover data categories ranging from maps, markers and genomic sequence through pedigree, phenotype and agroecology related. Possibilities for automated access to and utilization of resource data through Web services and workflows are described along with examples. Finally, the potential impacts of whole genome sequencing are noted, especially those affecting the range of data types and interface tools offered and the utilization of data exchange standards and ontologies.

1 Overview

The spectrum of *Gossypium* bioinformatics resources covers a broad range of life science areas and available data types of both genomic and phenotypic nature. For example, genetic map and diversity data are provided by genetics and evolutionary biology related resources and phenotypic data are provided by resources with focus areas that include crop science and biochemistry. In addition, multi-organism resources, such as those at the National Center for Biotechnology Information (NCBI), provide a substantial amount of *Gossypium* data that can be searched directly and are also available via other resources.

A.R. Gingle (✉)
Plant Genome Mapping Lab, University of Georgia
e-mail: agingle@uga.edu

Searching bioinformatics resources often involves multiple data types from multiple resources. In manual searches, this may be accommodated by cutting and pasting results from one resource's interface into that of another. However, this process rapidly becomes cumbersome as the volume of data increases. Many *Gossypium* resources provide navigational aids that automatically link to and provide needed data for other resources, relieving users of this tedious task. This provides a useful degree of data integration and search automation; however, more can be done to expand the range of questions that users can ask and the ease with which they can obtain integrated views of the resulting data. For example, a bioinformatics resource can make its functionality available over the Web to facilitate discovery and use by client tools that automate complex searches and data processing into workflows. There are emerging standards and tools for this higher level of integration and automation (Wilkinson and Links 2002; Stevens et al. 2003) and a brief summary of these technologies is included in a subsequent section of this chapter.

For typical searches, life science users have formulated questions and are approaching the Web for answers. As part of this process, users need to select entry resources from amongst the battery of those available. Of course, knowledge of a resource's available data types is essential for making this decision and, in addition, available data access methods may enter into the selection process. Accordingly, a survey of *Gossypium* bioinformatics resources was undertaken by the author and is the subject of the following section. It categorizes resources in terms of their life science focus areas, available data types and data access methods. A subsequent section employs these categories as a framework to illustrate search strategies as a function of data type.

The recommendations contained in this chapter are, of course, based on the *Gossypium* bioinformatics resources of the present. However, many of these resources are under active development and will likely expand their ranges of available data and, in some cases, data types as well. The advent of whole genome sequencing for multiple *Gossypium* species will lead to expansion of the volume of sequence data and the density of genomic descriptors that will drive the need for a higher level of data integration. Also, there will most likely be an increasing reliance on ontologies and related technologies that can facilitate the required data integration (Buccella et al. 2003; Lewis 2004; Hsing and Cherkasov 2006). Ontologies such as those for genome (GO), sequence (SO), expression (MAGE) and plant features (e.g., PO and TO) are expected to increase. The further development and implementation of data exchange methodologies and related standards will impact on the range of possible searches. These developments will be woven into the fabric of the latter sections of the chapter along with some extrapolation to aid in anticipating the future.

2 The Spectrum of *Gossypium* Informatics Resources

The primary goal of this chapter is to provide practical information for utilizing the array of *Gossypium* bioinformatics resources that are presently available via the Web. To establish the setting, a survey of *Gossypium* bioinformatics resources was undertaken by the author in early 2007. The survey included both *Gossypium* centric resources and those that provide data for *Gossypium* within a larger context. The author's battery of *Gossypium* related URLs, based on prior knowledge, was expanded through Web and publication searches. Resources were checked for available data access methods (e.g., data views, query tools, Web services, etc.) as illustrated in Fig. 3. The MOBY Central registry at the University of British Columbia (see table A.20 for URL) was surveyed for *Gossypium* related BioMOBY services. The database resources at the National Center for Biotechnology Information (NCBI) cover a diverse range of data types and, therefore, each database providing *Gossypium* data was treated as a unique resource in the survey.

2.1 Life Science Area and Available Data Types

Gossypium bioinformatics resources provide data and information from the perspectives of a wide range of life science areas. For example, the author's survey revealed resources with life science focuses that covered areas as diverse as crop science and structural genomics (Fig. 1). This survey included the full spectrum of *Gossypium* bioinformatics resources from those maintained by

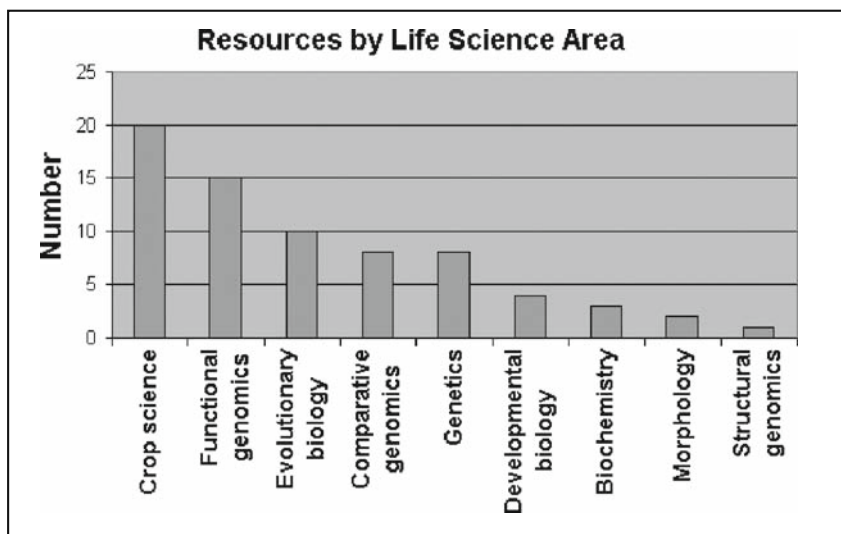


Fig. 1 A plot of *Gossypium* bioinformatics resources classified by life science area is shown

individual research labs or centers through those maintained by the U. S. Department of Agriculture (USDA) and onto large multi-organism resources with *Gossypium* data such as those at NCBI. The crop sciences are represented by the maximum number of resources with the majority of these devoted to cotton variety trial (CVT) performance data. The majority of the other life science areas is based in or have substantial overlaps with genomics (e.g., genetics as well as comparative and functional genomics) and is represented by approximately one half of the total for all areas. Thus genomics is well represented by the current spectrum of resources with the other areas often being of interest to genomics researchers.

Each of these life science areas has its own range of data types with their associated access, search and presentation features. The data types that were found to be available in the recent survey of bioinformatics resources are listed in table 1. They are grouped by categories that will be helpful to the example searches, appearing later in this chapter and a plot of bioinformatics resources grouped by these data types is shown in Fig. 2. As noted above, CVT performance data had the maximum representation. Genomics related data types such as genetic map and marker data are well represented (Rong et al. 2004; Frelichowski et al. 2006; Han et al. 2006) and with recent efforts the volume of available transcriptome and expression data have increased substantially (Udall et al. 2006; Udall et al. 2007). Also protein and biochemical pathway related data are represented.

Table 1 Available data types associated with *Gossypium* bioinformatics resources

Category	Data Type
Maps, markers and genomic sequence related	1. genetic map (individual loci and QTLs)
	2. genetic marker
	3. genomic sequence
	4. physical map and BAC related
Diversity and evolution related	5. genetic diversity
	6. molecular evolutionary
	7. taxonomic
Transcriptome related	8. expression/microarray
	9. expression/virtual
	10. transcriptome/EST
Proteome related	11. expression/protein
Phenotype related	12. protein domains, sequence or structure
	13. plant images
	14. biochemical pathways
Agriculture related	15. habitat and cultivation info
	16. pedigree
	17. performance trial
General information	18. bibliographic information
General bioinformatics tool related	19. sequence alignments

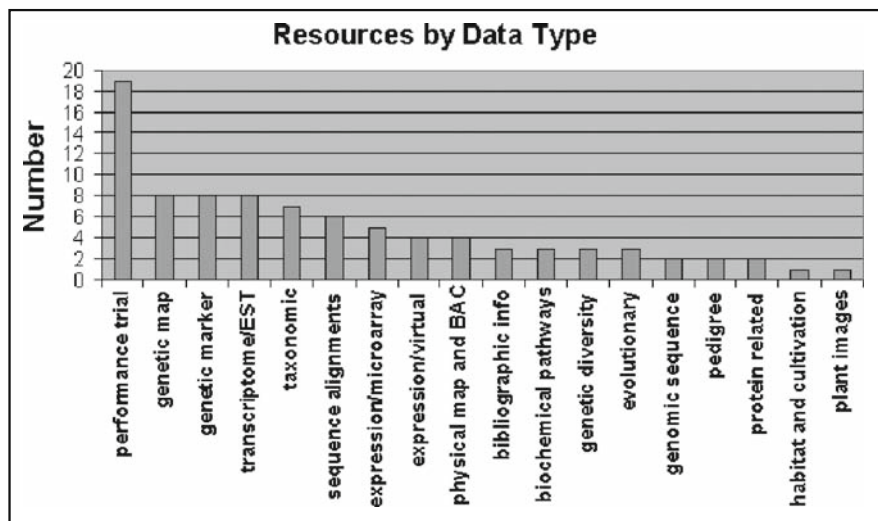


Fig. 2 A plot of *Gossypium* bioinformatics resources classified by primary data types is shown. Note that each resource providing a particular data type is counted even when they provide access to a common data set

Each of these data types has its own range of modalities for data display. For example, genetic map data is typically displayed in standard genetic linkage map format while genetic diversity data are typically displayed in phylogenetic tree or sequence alignment displays as exhibited by tools like Clustal (Chenna et al. 2003) and Web resources like the Cotton Diversity Database (Gingle et al. 2006) and PopSet at NCBI (see table A.15 for URL). Of course, Web displays for these and other data types have been developed, ranging in interactivity from static images to full featured interfaces that respond to mouse click events. Unlike data displays, search tools typically exhibit greater similarity across data types based on their common modes of entering search criteria (e.g., radio buttons as well as text, selection and check boxes). However, even they can exhibit data type distinctions especially when embedded in data display interfaces for convenience. In summary, a resource's available data types affect the kinds of data that can be queried, viewed and downloaded and, more importantly, they also impact on the range of search criteria that are available for asking questions.

2.2 *Methods of Data Storage and Accessibility are Important*

The manner in which data are stored and made accessible has a major impact on a resource's versatility with respect to searches and potential uses of its data by individual users and other resources. For example, data

that are stored and made available as a PDF or word processor document are only searchable via simple text searches. Record based flat files, such as those associated with spreadsheets (e.g., CSV), have greater utility since they can easily be uploaded into a relational database like Access or MySQL for subsequent querying. Data stored in a searchable database (e.g., a relational database or RDB) has even greater potential both in terms of the range of possible searches and ways in which it can be integrated with the data of other resources.

As an illustration of the impact of the modes of data storage and accessibility, consider the following basic genetic map related search that is common to most interactive displays for this type of data such as IntegratedMap (Yang et al. 2005) and CMap (see table A.20 for URL).

1. A user wishing to learn about all objects (e.g., BAC contig, EST, marker, etc.) associated with a particular genetic map region can specify that region using the entry resource's search tool or interface.
2. They obtain a list of genetic markers with related information stored at the resource.
3. They also obtain a list of BAC contigs with HTML links for related information stored at other resources.
4. The externally stored information contains physical map and BAC library data; however, this information is only available to the user after navigation and not via the entry resource where the search criteria are implemented.

In this example, the genetic map related data types not only specify types of data that are available for viewing or download, but also types that can be used as search criteria providing the data is stored in a searchable database with Web accessible search tools. The externally stored BAC library and physical map related information, however, is not available for search criteria since it is not stored in the resource's own database. For example, the search criteria could not be expanded to exclude all but those contigs containing BACs of a particular species. However, the expanded search criteria could have been employed if the library data were made available to the primary resource. Web services, software systems for resource-to-resource interaction and data exchange, can provide an effective way to achieve this by allowing the entry resource to automatically obtain the externally stored information and filter the returned results based on the selection criteria. The Wikipedia entry for Web services (see table A.20 for URL) provides a concise introduction to this technology. Web services are relatively uncommon amongst *Gossypium* resources; however, this is expected to change as developing standards and related tools gain community acceptance. They are described below in the context of automated data access (see section 4).

2.3 Use of Ontologies

The use of ontologies by *Gossypium* resources is, for the most part, confined to the genome ontology, GO (The Gene Ontology Consortium 2000 and 2004). Eleven resources were found to either provide GO term assignments or feature specific links to other resources of GO term assignments. In addition, MAGE (Whetzel et al. 2006) is employed for microarray expression data. These resources are listed in table 2 along with the ontology for which term assignments are available or used for data exchange.

Table 2 *Gossypium* Web resources that were found to provide ontology terms and/or employ ontologies for data exchange. Survey was taken by the author in February of 2007

Resource Name	Ontology
Cell Wall Navigator	GO
Comparative Evolutionary Genomics of Cotton	GO,MAGE
Cotton Functional Genomics Center	GO
Cotton Marker Database	GO
DFCI Cotton Gene Index	GO
NCBI (dbGSS)	GO
NCBI (Entrez Gene)	GO
NCBI (GEO)	MAGE
NCBI (Popset)	GO
NCBI (Taxonomy)	GO
NCBI (UniGene)	GO
UniProt	GO

2.4 Available Data Access Methods

Given their importance for searches and data integration, *Gossypium* bioinformatics resources were also grouped by some common data access methods as part of the survey. The number of bioinformatics resources classified in this way is plotted (Fig. 3) for Web accessible data views, Web accessible query tools, PDF and other document availability, bulk data flat file availability, and Web and other services. As one might expect, static Web page data views are the most common since they provide an easy way to present data with minimal bioinformatics overhead. Resources providing PDF or other document downloads are especially common for crop sciences resources focused on CVT performance data. Web accessible query/search tools based on relational data (stored in a RDB) are quite common as are resources providing bulk data downloads. As noted above, Web services are relatively uncommon. While eleven resources provide data in this fashion with published and/or registered formats, seven are components of NCBI.

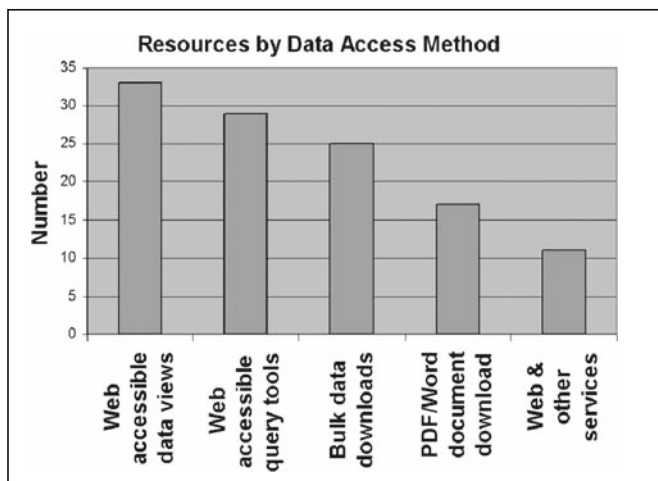


Fig. 3 A plot of *Gossypium* bioinformatics resources classified by data access features is shown

3 Searching for *Gossypium* Data and Information

Optimal search methods and strategies for *Gossypium* resources are dependent on the available data formats and search tools that are provided by the relevant resources. These depend on the primary data types that are involved with a particular search since many of the resource data formats, access methodologies and search interfaces have developed around data type specific requirements. In this section, example searches highlight these idiosyncrasies for many of the data categories and types as listed in table 1. They are illustrated through workflow diagrams (Figs. 4–10) that highlight the commonalities where they exist. Some resource specific information and methodology are also included, especially for the more integrative searches that cover a broader range of data types.

3.1 Maps, Markers and Genomic Sequence

These searches involve map related features and, since only genetic maps are present for *Gossypium* (Rong et al., 2005; Frelichowski et al. 2006; Han et al. 2006), they are mostly focused on genetic loci as well as related probes and traits. However, both physical map development (L. Lin, G. Pierce, J. Bowers, A. H. Paterson, unpublished results) and whole genome sequencing (Joint Genome Institute Community Sequencing Program) are under way for *Gossypium*. Searches that involve BACs, genomic sequences and STSs, which are some of the components and features of these developing genomic landscapes, are also included.

Searches by genetic map location or feature are illustrated in the diagrams of Fig. 4. Three resources providing *Gossypium* genetic map data are included. Both CottonDB (CDB) and TropGENE-DB (TGDB) facilitate these searches

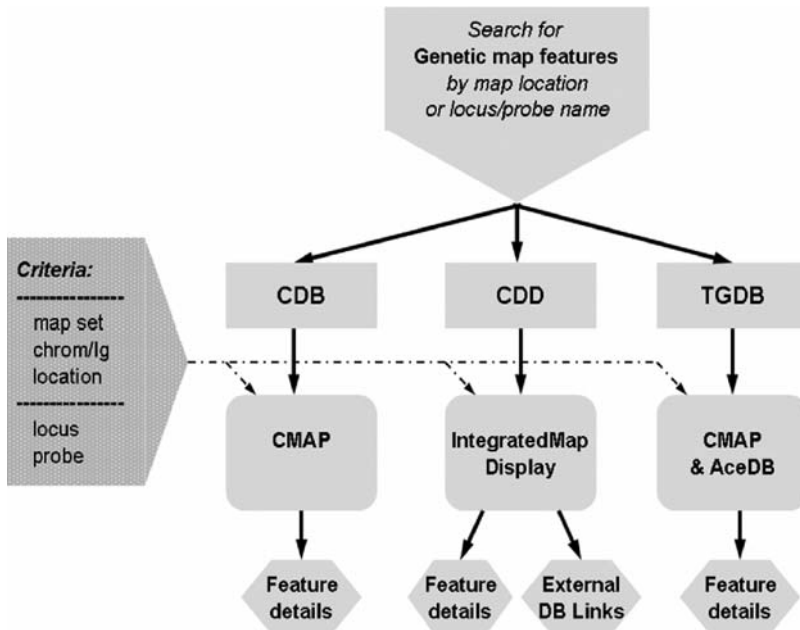


Fig. 4 Searches for genetic map features are illustrated in the figure. Three resources that provide *Gossypium* genetic map data were sampled by the author and are included. They are CottonDB (CDB), the Cotton Diversity Database (CDD) and TropGENE-DB (TGDB). Possible criteria for searches are map set, chromosome or linkage group, location and locus or probe name. Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes

through implementations of the CMap display tool that, in these settings, provides single and comparative genetic map views of the available maps. At the time of this writing, CDB was found to provide data from four maps based on *G. hirsutum* and *G. barbadense* x *G. hirsutum* crosses and TGDB was found to provide data from five *G. hirsutum* based maps. TGDB also provides marker data via its AceDB implementation (Ruiz et al. 2004). The Cotton Diversity Database (CDD) focuses on the high density *Gossypium* map generated by Rong et al. (Rong et al., 2005) and provides genetic map views and search capabilities via the IntegratedMap tool (Gingle et al. 2006; Yang et al. 2005). All of these resources provide both genetic map views and information about the mapped features that result from the searches. CDD also provides links to NCBI for feature related nucleotide data and PGML when developing physical map related data is available.

Searches for other map related components and features are illustrated in the diagrams of Fig. 5. The three searches involve distinct data types and resources and are grouped into a single figure for the sake of compactness. At NCBI, searches for genomic sequences and sequence tag sites (STS), by Id or accession,

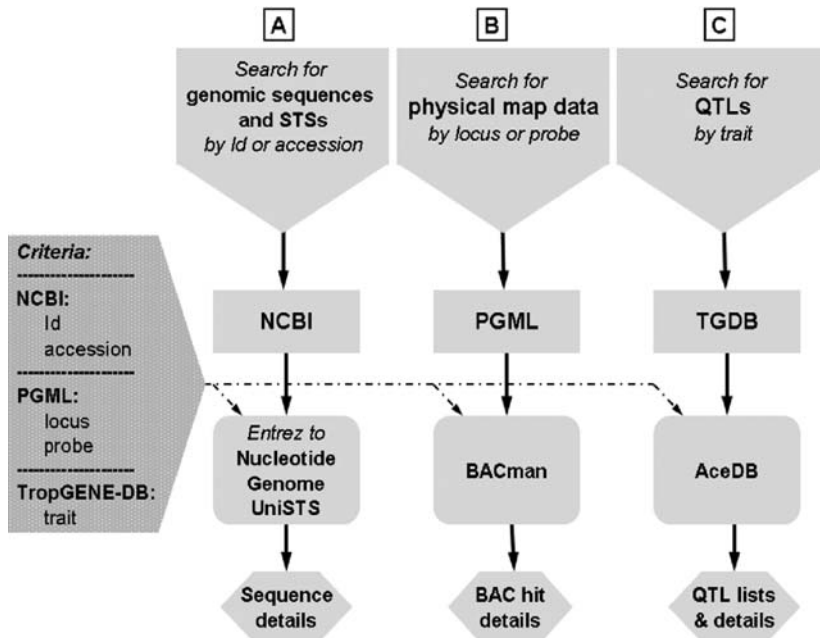


Fig. 5 Searches for genomic, physical map and QTL related data are illustrated in the figure. Three resources that provide related *Gossypium* data were sampled by the author and are included. They are the National Center for Biotechnology Information (NCBI), the Plant Genome Mapping Laboratory (PGML) and TropGENE-DB (TGDB). At NCBI, searches for genomic sequences and sequence tag sites (STS), by Id or accession, are directed to the relevant database via their Entrez search tool (workflow A). At PGML, searches for physical map data, by locus or probe name, are facilitated by their BACman resource (workflow B). At TropGENE-DB, searches for QTLs, by trait, are facilitated by their AceDB based resource for this data type (workflow C). Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes

are directed to the relevant database via their Entrez search tool. Their search tool is based on text strings that can optionally include Boolean operators (Benson et al. 2006). Record counts for each of their databases are displayed as links that ultimately lead to their data display tools. At PGML, searches for the developing physical map data, by locus or probe name, are facilitated by the BACman (Estill et al. 2003) resource, which provides probe-BAC hybridization data for BAC libraries derived from *G. barbadense*, *G. hirsutum* and *G. raimondii*. At TropGENE-DB, searches for QTLs, by trait, are facilitated by their AceDB based resource for this data type. Their search tool contains selector boxes for trait name, linkage group and correlation threshold and returns associated QTLs with related data (Ruiz et al. 2004). A recent effort at PGML is providing search capabilities, via CMap, for their QTL database, containing 432 loci related to a range of development, morphology, stress tolerance and yield traits (Rong et al. in press).

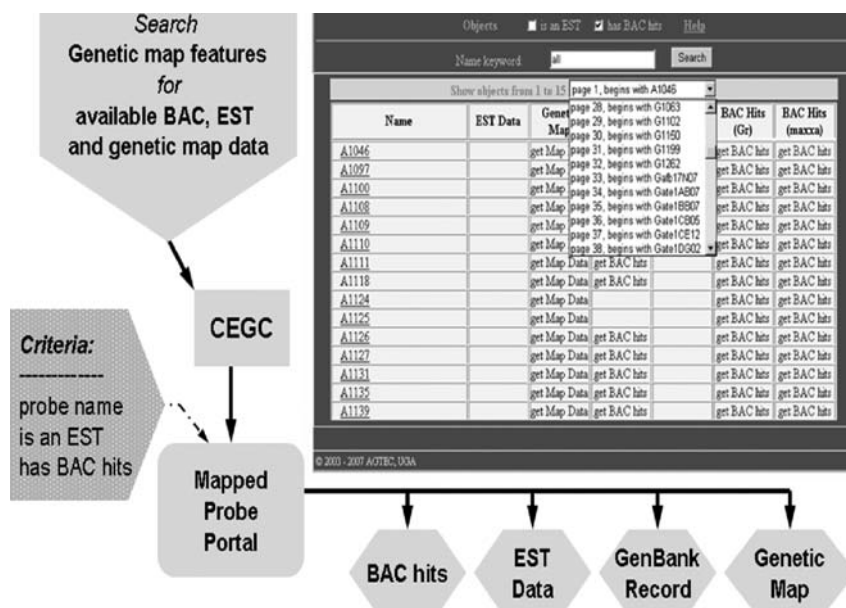


Fig. 6 A workflow and tool for searching genetic map features in an integrated fashion is illustrated in the figure. The tool, Mapped Probe Portal, is one component of the suite of Web interfaces that are available at the Comparative Evolutionary Genomics of Cotton (CEGC) resource. It facilitates searches for all resource data associated with genetically mapped probes based on criteria that include probe name, hits to *Gossypium* BACs and whether the probe is an EST. The interface displays links that allow users to navigate to other resources and tools for the range of available data types, which are illustrated by the hexagons in the workflow diagram. The external resources and data type associations are described in the text. Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes. See table A.20 for Web tool URL.

A workflow and tool for searching genetic map features in an integrated fashion is illustrated in Fig. 6. The tool, Mapped Probe Portal, is one component of the suite of Web interfaces that are available at the Comparative Evolutionary Genomics of Cotton (CEGC) resource. It facilitates searches for all resource data associated with genetically mapped probes based on criteria that include probe name, hits to *Gossypium* BACs and whether the probe is an EST. The interface displays links that allow users to navigate to other resources and tools for the range of available data types. BAC hit data is accessed via links to PGML's BACman (Estill et al. 2003) Web tool, mapped EST data via links to the AGCol PAVE Web tool for EST assemblies (see table A.17 for URL), GenBank records via links to NCBI, and related genetic map views via links to the CDD IntegratedMap Web tool (Yang et al. 2005).

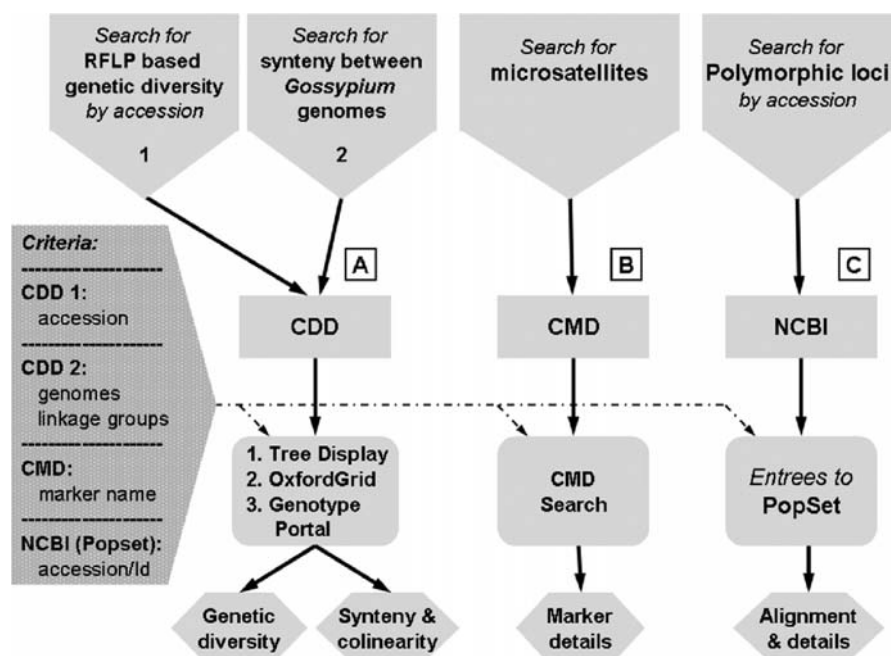


Fig. 7 Genetic diversity related searches are illustrated in the figure. Three resources that provide related *Gossypium* data were sampled by the author and are included. They are the Cotton Diversity Database (CDD), the Cotton Marker Database (CMD) and the National Center for Biotechnology Information (NCBI). At CDD, searches for genetically diverse or similar accessions typically begin with an accession of interest and are facilitated by their Genotype Portal search tool and Phylogenetic Tree display (workflow A1). Comparative genetic map based synteny/colinearity searches are facilitated by their Genotype Portal search tool and OxfordGrid display for dot plot type display of comparative genetic map data (workflow A2). At CMD, searches for *Gossypium* microsatellites are facilitated by their search tool (workflow B). At NCBI, searches for polymorphic loci are routed to their PopSet database display tools (workflow C). Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes

3.2 Diversity and Evolution

The manifestations of genetic diversity can range from an individual polymorphic locus through interruptions in microsynteny or the lack of colinearity between duplicated chromosomes. Therefore, the example searches involve a diverse range of criteria and data types as illustrated in the workflows of Fig. 7. They involve three resources that provide diversity related data for *Gossypium* and were sampled by the author. They are the Cotton Diversity Database (CDD), providing comparative map and RFLP based genetic diversity data; the Cotton Marker Database (CMD), providing SSR related data; and PopSet at the National Center for Biotechnology Information (NCBI), providing in/del

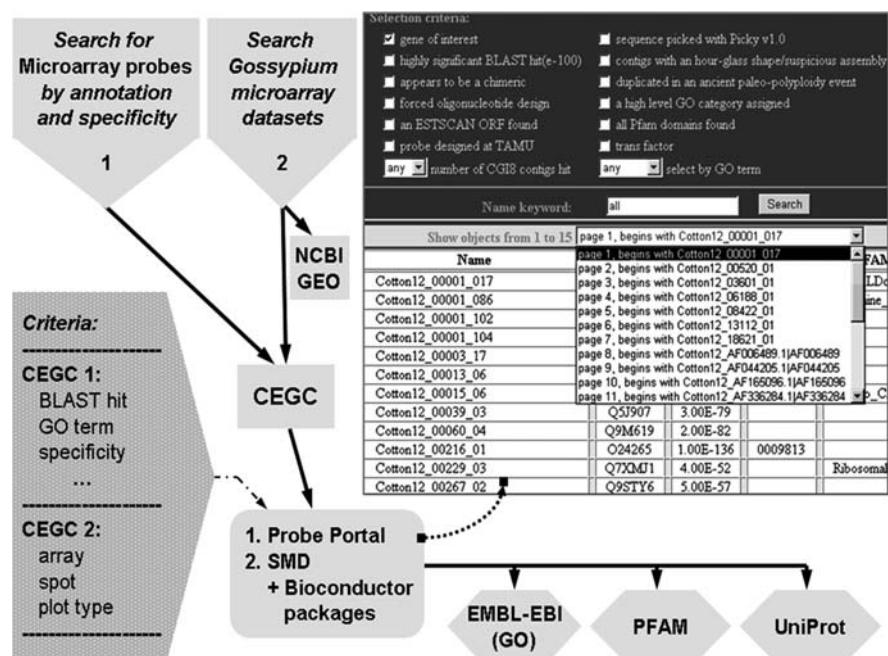


Fig. 8 A workflow and tool for searching microarray expression data is illustrated in the figure. The tools, Array Probe Portal (1) and an implementation of the Stanford Microarray database (SMD, 2) with Bioconductor additions, are components of the suite of Web interfaces that are available at the Comparative Evolutionary Genomics of Cotton (CEGC) resource. They facilitate searches for microarray probe annotation and specificity data as well as microarray datasets that have been generated by the related project. The Array Probe Portal interface displays links that allow users to navigate to other resources and tools for the range of probe related data, which are illustrated by the hexagons in the workflow diagram. The external resources and data type associations are described in the text. The SMD database provides access to array datasets, data plots and platform information. This information is also available from GEO at NCBI (see related text for more info). Navigation paths are indicated by solid arrows, search criteria entry by dashed lines

and SNP data for polymorphic loci. Each resource has its unique search and display formats as highlighted below.

At CDD, searches for genetically diverse or similar accessions (workflow A1) typically begin with an accession of interest and are facilitated by their Genotype Portal search tool and Phylogenetic Tree display (Gingle et al. 2006). Search criteria are entered by selecting a particular accession or sub-tree via mouse clicks on the interactive display and data is displayed in a phylogenetic tree format. Comparative genetic map based synteny/colinearity searches (workflow A2) are facilitated by their Genotype Portal search tool and Oxford-Grid display for dot plot type display of comparative genetic map data (Gingle

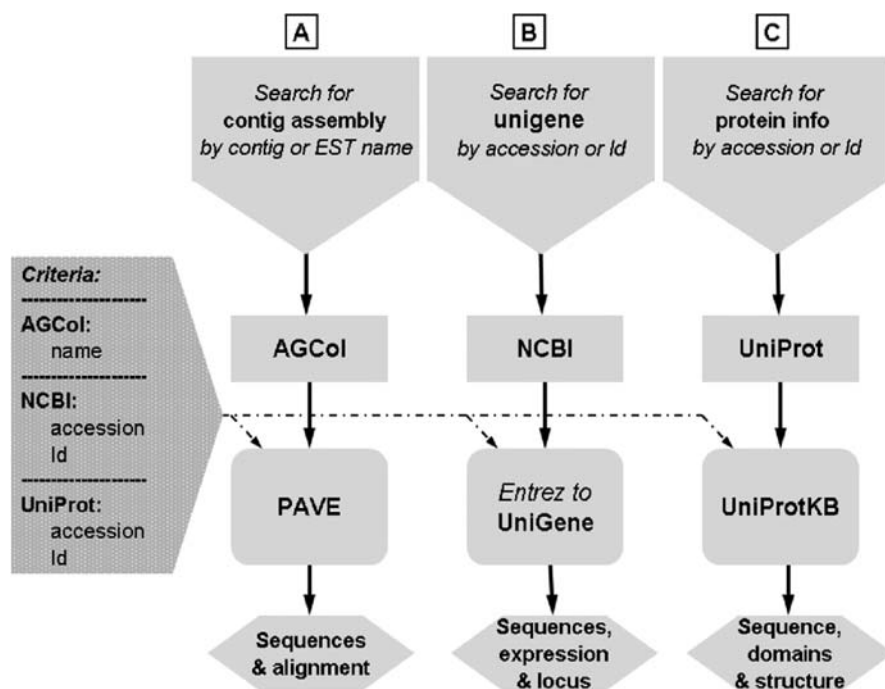


Fig. 9 Searches for transcriptome and proteome related data are illustrated in the figure. Three resources that provide related *Gossypium* data were sampled by the author and are included. They are Arizona Genomics Computational Lab (AGCol), the National Center for Biotechnology Information (NCBI) and the Universal Protein Resource (UniProt). At AGCol, searches for EST contigs and related information by contig or member EST name are facilitated by their PAVE tool for assembly viewing (workflow A). At NCBI, searches for unigene EST clusters and related sequence, locus and expression data, by Id or accession, are directed to the relevant database via their Entrez search tool (workflow B). At UniProt, searches for available protein sequence, domain and structural data, by Id or accession, are facilitated by their UniProtKB search tool (workflow C). Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes

et al. 2006; Yang et al. 2005). Search criteria are entered by selecting genome and linkage group/chromosome combinations and data is displayed in an interactive dot plot format that allows the selection of syntenic locations for further detailed information. At CMD, searches for *Gossypium* microsatellites (workflow B) are facilitated by their search tool (Blenda et al. 2006). Search criteria are entered by text or selection of SSR(s) from lists and data is displayed in html/text format. At NCBI, searches for polymorphic loci (workflow C) are routed to their PopSet database display tools (see table A.15 for URL). Search criteria are entered by text or selection from lists and data is displayed in a combination of graphic and html/text formats.

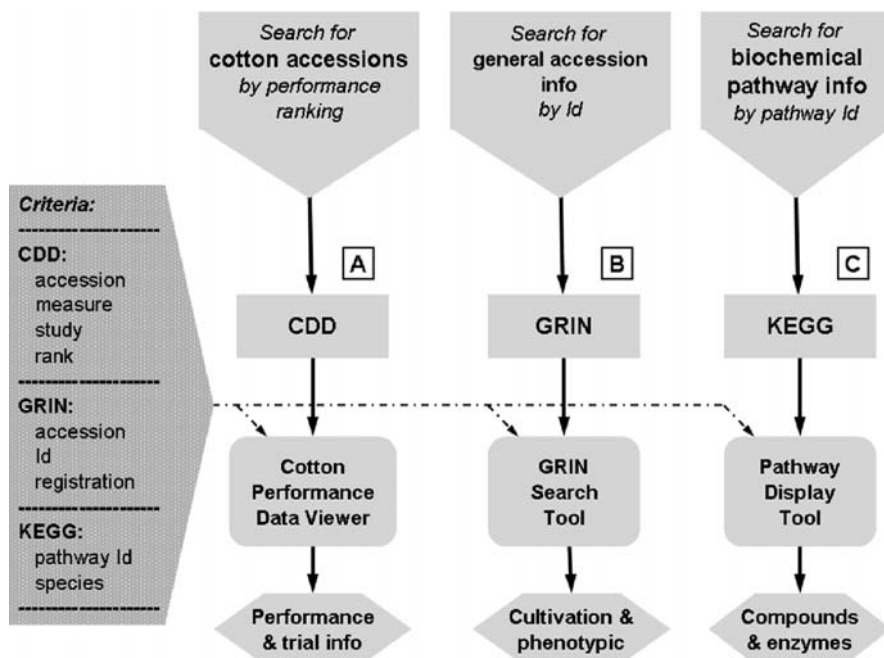


Fig. 10 Searches for biochemical, phenotypic and performance trial related data are illustrated in the figure. Three resources that provide related *Gossypium* data were sampled by the author and are included. They are the Cotton Diversity Database (CDD), Germplasm Resources Information Network (GRIN) and Kyoto Encyclopedia of Genes and Genomes (KEGG). At CDD, searches for accessions and related study information by performance measure ranking are facilitated by their performance data viewer tool (workflow A). At GRIN, searches for cultivation, phenotypic and general accession related information, by Id, are facilitated by their search tool (workflow B). At KEGG, searches for biochemical pathway information are facilitated by their pathway display tool (workflow C). Navigation paths are indicated by solid arrows, search criteria entry by dashed lines and data availability highlights are indicated in the hexagonal boxes

3.3 Transcriptome and Proteome

Gossypium transcriptome assets are quite substantial. For example, at least four spotted microarray platforms (Arpat et al. 2004; Lee et al. 2006; Udall et al. 2007) have been generated to date. An updated version of the cotton oligonucleotide microarray that contains more than 21,000 probes assembled from more than 200,000 of the available *Gossypium* ESTs, representing 52 cDNA libraries (Udall et al. 2007). See table A.20 for related URL. In addition, Affymetrix has recently developed a cotton genome array comprised of 23,977 probe sets representing 21,854 *Gossypium* transcripts with information on this array, available via their Web site (see table A.20 for URL). Overall, four *Gossypium* species (*G. arboreum*, *G. barbadense*, *G. hirsutum* and

G. raimondii) are represented by these arrays. As one might expect, *Gossypium* proteomic resources are not as substantial; however, they do exist and one example search is devoted to data of that category. The example searches involve a diverse range of criteria and data types as illustrated in the workflows of Figs. 8–9. They involve four resources, The Arizona Genomics Computational laboratory (AGCol), the Comparative Evolutionary Genomics of Cotton (CEGC), UniGene at the National Center for Biotechnology Information (NCBI) and the Universal Protein Resource (UniProt). Three of these resources are providing transcriptome related data with the other providing primarily proteome related data in these search examples.

The examples begin with a workflow and tool for searching microarray expression data, which are illustrated in the Fig. 8. The examples include searches for microarray probe annotation and specificity data as well as microarray datasets that have been generated by the related CEGC project using their cotton oligonucleotide microarray (Udall et al. 2007). The workflows employ two informatics tools, the Array Probe Portal (see table A.20 for URL) and an implementation of the Stanford Microarray database (SMD; Ball et al. 2005) with Bioconductor (Gentleman et al. 2005) additions. The tools are components of the suite of Web interfaces that are available at the CEGC resource. The Array Probe Portal interface displays links that allow users to navigate to other, including external, resources for the full range of available probe related data. Possible search criteria include assigned GO term, PFAM domain assignment, specificity in terms of vmatch (see table A.20 for URL) hits to cotton genome index (see table A.17 for URL) sequences, etc. The linked external resources are European Molecular Biology Laboratory (EMBL-EBI) for GO term related information, Protein Families database at the Sanger Institute (PFAM) for assigned protein domain information and the Universal Protein Resource (UniProt) for protein information related by significant BLAST hits. The SMD database provides access to array datasets, data plots and platform information. This information is also available from the Genome Expression Omnibus (GEO) at NCBI (Barrett et al. 2005; Edgar and Barrett 2006) and can be searched by the associated GEO microarray platform and dataset accession Ids.

Other searches for transcriptome and protein related data are illustrated in Fig 9. Three resources that provide related *Gossypium* data were sampled by the author and include AGCol, NCBI UniGene and UniProt. At AGCol, searches for EST contigs and related information by contig or member EST name are facilitated by their PAVE tool (see table A.17 for URL) for assembly viewing (workflow A). Search criteria are entered via a combination of text fields and list selections and data is displayed in a combination of graphic and html/text formats. At NCBI, searches for unigene EST clusters and related sequence, locus and expression data, by Id or accession, are directed to the relevant database via their Entrez search tool (workflow B). Search criteria are entered by text or selection from lists and data is displayed in a combination of graphic and html/text formats. At UniProt, searches for available protein sequence, domain and

structural data, by Id or accession, are facilitated by their UniProtKB (Schneider et al. 2005) search tool (workflow C). Search criteria are entered by a combination of text fields and list selections and data is displayed in text/html format with links to other resources (e.g., NCBI Taxonomy and TrEMBL).

3.4 *Phenotype*

Resources providing *Gossypium* phenotypic data range in focus from performance trial results to biochemical phenotypes such as those exhibited by enzymes associated with biochemical pathways. Therefore, the search examples, illustrated in Fig.10, are quite varied. The largest portion of phenotype related resources were found to be devoted to performance trial data (see Fig. 2). However, most provide trial data in formats that are not easily searched. One exception is the Cotton Diversity Database (CDD), which provides search tools for trial data stored in a relational format (Gingle et al. 2006). It is illustrated in workflow A of Fig. 10. At CDD, searches for accessions and related study information by performance measure ranking are facilitated by their performance data viewer tool. Search criteria are entered by selecting a particular accession or performance ranking based on any of the common measures (e.g., fiber strength, micronaire, etc.) and data is displayed in a graphical format with links to detailed trial information. A second search example involves information of a more general nature that is available at the Germplasm Resources Information Network (GRIN) and is illustrated in workflow B of Fig. 10. At GRIN, searches for cultivation, phenotypic and general accession related information, by Id, are facilitated by their search tool (see table A.15 for URL). Search criteria are entered by a combination of text entry and list selections and data is displayed in html/text format.

3.5 *Biochemical*

While biochemical data can be considered as phenotypic in nature, the Kyoto Encyclopedia of Genes and Genomes (KEGG) resource is quite distinct from others offering agricultural related phenotypic data. KEGG contains a database that provides biochemical pathway information and an example search is illustrated in workflow C of Fig. 10. At KEGG, searches for biochemical pathway information are facilitated by their pathway display tool (Ogata et al. 1999). Search criteria are entered by a combination of text entry and list

selections and data is displayed in a combination of html/text and interactive image map formats.

4 Automated Access to Resource Data

Automated access to *Gossypium* resource data can enable search tools that provide users with an expanded set of search options, as noted in section two. This type of data integration can be facilitated via Web services, which are based on software systems that allow resource computers to interact and exchange data. Web services can be accessed in a coordinated fashion by workflows which can combine the information from multiple resources to meet the needs of bioinformatics data processing and data retrieval. Their potential benefits are significant; so much so that tools like Remora (Carrere and Gouzy 2006) and Taverna (Oinn et al. 2004) have been developed for bioinformatics workflow creation. The growing importance of Web services is also underscored by the development of the Bioinformatics communication protocols, BioMoby (Wilkinson and Links 2002) and mygrid (Stevens et al. 2003), that facilitate automated discovery of available services. Thus, even though the current use of Web services by *Gossypium* informatics resources is limited, their growing importance to Bioinformatics, in general, will likely be reflected in the future developments of *Gossypium* resources. Therefore, some Web service and workflow basics are included in the following sections along with examples.

4.1 Web Service Basics

Web services are typically interfaces for computer programs that can be accessed via the internet. They allow one computer to remotely call a procedure or program that is executed on another computer, which hosts the procedure. When applied to automated data access, the remotely executed procedures typically involve retrieving, formatting and returning requested data such as the BAC library data in the example of section two. When used for automating searches, the search criteria are typically formatted as parameters that are sent to the relevant service. There are, of course, specifications for requesting and providing these services and they are defined by the W3C standard (see table A.20 for URL). Three of these specifications, Simple Object Access Protocol (SOAP), Web Service Description Language (WSDL) and the Universal Description, Discovery, and Integration (UDDI) registry service, are central to Web service operation. SOAP specifies an extensible markup language (XML) based message format that encapsulates requests for data and/or data processing. It specifies the way that Web services encapsulate requests and data

that are bound to internet protocols like HTTP, the communications protocol that enables Web browsing for example. WSDL specifies the language for describing Web services (see table A.20 for URL). A WSDL file provides needed information for interfacing with its associated Web service and can be Web published for use by software that accesses the Web service. UDDI (see table A.20 for URL) is a protocol for publishing and discovering this information or metadata about Web services. BioMOBY has further developed Web service discovery for the life sciences through the development and implementation of registries and ontologies for related Web services (Wilkinson and Links 2002). These Web service standards and components are illustrated in a few examples.

4.2 Gossypium Data from Web Services

While the use of Web services by *Gossypium* informatics resources is limited, they are available as described below and illustrated in Fig. 11. Both the CDD and CEGC resources provide genetic map related Web services with published descriptions and WSDL specifications. The CDD/CEGC Web services include methods that range from those for obtaining a listing of the available genetic maps, linkage groups and other related information to others for obtaining features by map region. They can be accessed directly from the resource or via client software applications like Taverna (Oinn et al. 2004), which facilitates combining Web services into workflows. When using Taverna, these services can be accessed either by manually adding their WSDL specification URL (see Fig. 11) to the services selection panel as a WSDL scavenger or, in a more automated fashion, by including their WSDL description URL in the Taverna mygrid.properties file. The multi-taxa resources, NCBI and KEGG, also provide Web services with published descriptions and were found to be included as default services in recent versions of Taverna. At NCBI, the eUtils Web services include methods that provide automated programmatic access to their Entrez utilities and they provide help documentation for their Web services (see table A.20 for URL). An example is their EGQuery method that returns information on the number of records in each of their databases that match the search criteria, which can include general terms like “*Gossypium*” or specific accession Ids. At KEGG, Web services include methods that provide access to biochemical pathway information based on search criteria that include related genes, enzymes, compounds, reactions, etc. (Kawashima et al. 2003). For example, their `get_pathways_by_genes` method searches pathways by gene Ids. Also, TropGeneDB offers genetic diversity and map related BioMOBY services that are also accessible via Taverna.

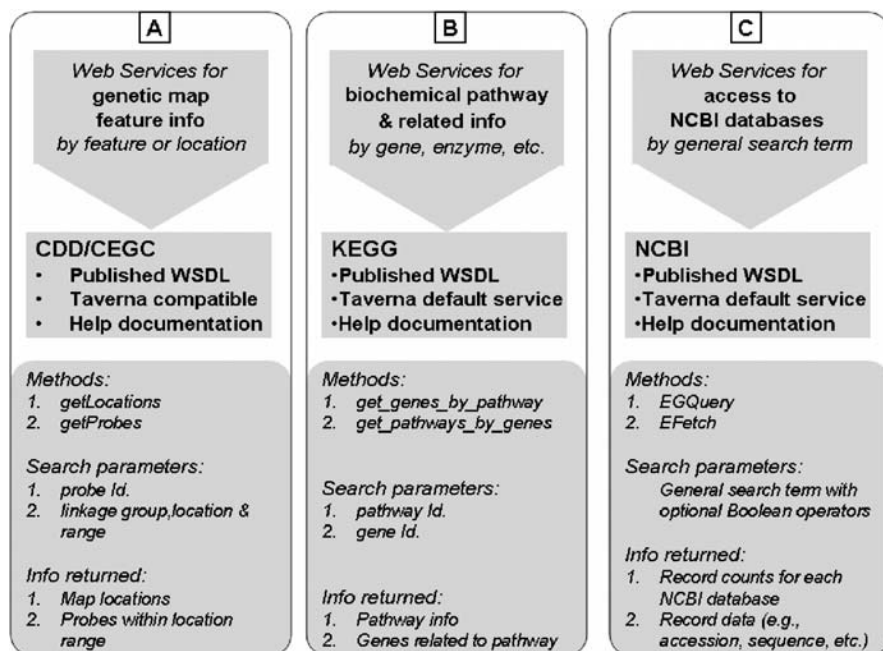


Fig. 11 Web services that provide *Gossypium* data are illustrated in the figure. Both the CDD and CEGC resources provide services for their genetic map data (A). KEGG provides services for their biochemical pathway information (B) and NCBI provides services for accessing their Entrez utilities (C). The services are provided with published descriptions and WSDL specifications that can be accessed via client software applications like Taverna. These Web services are also included in the example workflow of section 4.3

4.3 Workflows

As noted above, workflows are an important application area for Web services and workflow design tools like Taverna provide work benches that allow users to combine Web services for more integrated data processing and searches. Workflows can, of course, be implemented in programming languages like Java, Perl and Python. However, Taverna offers graphical workflow design features that may be helpful to the non programmer who is trying to learn about Web services and how they can integrate into their bioinformatics data processing environment. So, the following example will illustrate a simple workflow, employing CDD/CEGC and NCBI Web services and set in the Taverna work bench environment.

The example workflow is designed to process a GenBank accession Id with both CDD and NCBI Web services, yielding any associated *Gossypium* genetic map locations and statistics on any database records at NCBI (Fig. 12). The CDD Web service, *getProbeLocations*, returns the *Gossypium* genetic map coordinates for any accession that appears on the high density map of Rong et al. (2004). The NCBI Web service, *EGQuery*, returns

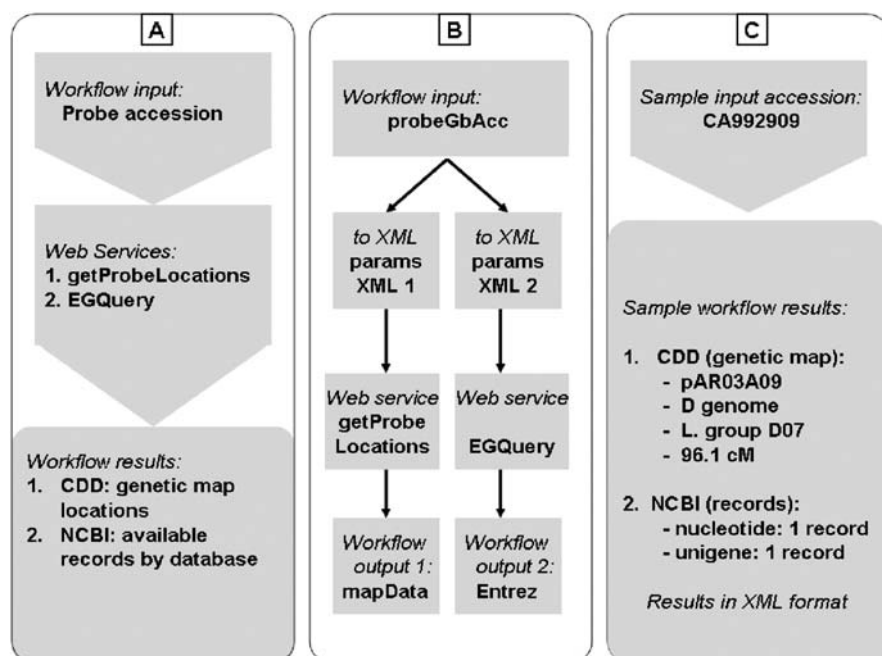


Fig. 12 A simple workflow is illustrated in the figure. The workflow processes a GenBank accession Id with both CDD and NCBI Web services, yielding any associated *Gossypium* genetic map locations and statistics on any database records at NCBI (A). The workflow diagram, as implemented in the Taverna environment, contains components for the plain text input, XML formatting to accommodate the Web Services at CDD (getProbeLocations) and NCBI (EGQuery), and data outputs/results in XML format (B). Sample input and results are shown in panel C. Note, the results are actually returned in XML format. They are summarized as a list in the figure for the sake of compactness

statistics on any database records at NCBI. The input, Web services and results are shown in Fig. 12A and a workflow diagram is shown in Fig. 12B. The workflow consists of the accession input (probeGbAcc), the two Web services, some components for XML formatting and the outputs for related map data and NCBI record statistics. An example input parameter and output results are shown in the figure (Fig. 12C). Note, the results are actually returned in XML format. They are, however, summarized as a list in the figure for the sake of compactness. The example, by design, is a very simple workflow; however, it could be expanded through the inclusion of other web services such as EFetch for retrieving NCBI data. Also, workflows can be designed around the KEGG Web service (e.g., get_pathways_by_genes) for biochemical pathway information and incorporate any of the BioMOBY services (Kawas et al. 2006) such as those at TropGeneDB (Ruiz et al. 2004). More complex *Gossypium* related workflows will become possible as the range of available Web services grow.

5 The Impact of Whole Genome Sequencing

Whole genome sequencing of *Gossypium raimondii* that will achieve 0.5x coverage is under way as a result of an award from the Joint Genome Initiative, JGI, (A draft sequence of the simplest *Gossypium* (cotton) genome, PI: A. Paterson). This will likely lead to additional genome sequencing efforts, most likely targeted to increasing coverage in *G. raimondii* and eventually other *Gossypium* species. These efforts will have substantial impacts on the community of bioinformatics resources offering *Gossypium* data. At the most basic level, the genomic sequence will begin to form a reference for annotations and linkages to the genetic (Rong et al., 2005) and developing physical (L. Lin, G. Pierce, J. Bowers, A. H. Paterson, unpublished results) maps for *Gossypium*. Web interfaces for genome browsing, like GBrowse (Stein et al., 2002) or VISTA (Shah et al. 2004), will add to the existing battery of tools for *Gossypium* data. As part of their participation in *Gossypium* genome sequencing, JGI's Web resources will increasingly provide *Gossypium* genomic data as they are for other plants like Poplar and Sorghum. Accordingly, a brief description of their resources appears below. Also, as the genome coverage increases, annotation systems such as the Distributed Annotation System or DAS (Dowell et al. 2001), data exchange standards such as GFF3 (see table A.20 for URL) and related ontologies such as SO (Eilbeck et al. 2005) will play an increasing role in the *Gossypium* bioinformatics resource landscape.

The primary JGI Web tool is their genome viewer suite which displays tracks that represent the scaffold sequence, gene models and other relevant sequences. In addition to general text searches, their search tools facilitate sequence alignment, track feature and gene model based searches. Menu options provide on-line BLAST, sequence data downloads and GO, KEGG and KOG (Tatusov et al. 2003) related information. Their resource for poplar, *Populus trichocarpa*, is accessible (see table A.20 for URL) and offers a good example of the functionality that will likely be available for *Gossypium* at some future time.

Appendix A

A.1 *Gossypium* Web Resource Survey: Table of Resources

This appendix contains tables of Web resources that provide *Gossypium* data in each of the listed categories. A final table contains other URLs of projects and resources that are discussed in the text. Tables for each data category list resource name or abbreviation and URL. Resources providing multiple categories of data appear in multiple tables.

A.11 Maps, Markers and Genomic Sequence

Name and URL

Arizona Genomics Institute

<http://www.genome.arizona.edu/genome/cotton.html>

Analysis of Early Events in Cotton Fiber Development

<http://www.cottongenomics.org>

Comparative Evolutionary Genomics of Cotton (CEGC)

<http://cottonrevolution.info>

CottonDB

<http://cottondb.org>

Cotton Diversity Database

<http://cotton.agtec.uga.edu>

Cotton Marker Database

<http://www.cottonmarker.org>

NCBI (dbGSS)

<http://www.ncbi.nlm.nih.gov/dbGSS/>

The Computational Biology and Functional Genomics Laboratory (CGI)

<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cotton>

Plant Genome Mapping Lab (Cotton)

<http://www.plantgenome.uga.edu/cotton/CottonDBFrames.htm>

Plant Genome Mapping Lab (syteny and comparative QTL data)

<http://www.plantgenome.uga.edu/cmap>

TropGENE-DB (Cotton)

<http://tropgenedb.cirad.fr/en/cotton.html>

A.12 Plant images

Name and URL

DFT Digital Library (*Gossypium*)http://www.cSDL.tamu.edu/FLORA/cgi/gallery_query?q=gossypium

A.13 Crop Science Related

Name and URL

Alabama CVT

<http://www.ag.auburn.edu/agrn/alabamavarietytesting/>

Arizona CVT

<http://ag.arizona.edu/crop/cotton/varieties/varietytrials.html>

Arkansas CVT

<http://comp.uark.edu/~avrtest/index.php?cotton&content>

California CVT

<http://cottoninfo.ucdavis.edu>

Cotton Diversity Database

<http://cotton.agtec.uga.edu>

Florida CVT

http://www.extension.org/pages/Florida_Cotton_Varieties

Georgia CVT

<http://commodities.caes.uga.edu/fieldcrops/cotton/varieties/var2006.htm>GRIN

A.13 (continued)

Name and URL
http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl
Kansas CVT
http://kscroptests.agron.ksu.edu/03/03cotton/3n-test.asp
Mississippi CVT
http://msucares.com/pubs/crops3.html#cotton
Missouri CVT
http://aes.missouri.edu/delta/cotton/index.stm
National CVT
http://www.ars.usda.gov/Business/Business.htm?docid=4357&modecode=64-02-15-00&page=2
New Mexico CVT
http://www.cahe.nmsu.edu/pubs/variety_trials/welcome.html
North Carolina CVT
http://www.ovt.ncsu.edu/ovt.asp?fmtinfo=1,450,150
Oklahoma CVT
http://www.osu.altus.ok.us/Extension%20Cotton%20Annual%20report.html
Plants Database
http://plants.usda.gov
South Carolina CVT
http://www.clemson.edu/agronomy/VT/Cotton/cotton.htm
Tennessee CVT
http://www.utextension.utk.edu/fieldCrops/cotton/varieties.htm
Texas CVT
http://varietytesting.tamu.edu
Virginia CVT
http://www.ext.vt.edu/cgi-bin/WebObjects/Docs.woa/wa/getcat?cat=ir-cg-cr-co

A.14 Proteome Related

Name and URL
Cell Wall Navigator
http://bioweb.ucr.edu/Cellwall/index.pl
UniProt
http://www.expasy.uniprot.org

A.15 Diversity and Evolution

Name and URL
Botanical Name Portal
http://www.anbg.gov.au/ibis/speciesLinks.html
Cotton Diversity Database
http://cotton.agtec.uga.edu
Global Biodiversity Information Facility (GBIF)
http://www.gbif.org
Genetic Resource Information Network (GRIN)
http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl

A.15 (continued)

Name and URL

Integrated Taxonomic Information System (ITIS)

<http://www.itis.gov>

NCBI (PopSet)

<http://eutils.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Popset>

NCBI (Taxonomy)

<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>

Plant Genome Mapping Lab (syteny and comparative QTL data)

<http://www.plantgenome.uga.edu/cmap>

Plants Database

<http://plants.usda.gov>

The Wendel lab

<http://www.eeob.iastate.edu/faculty/WendelJ/home.htm>

A.16 General Information

Name and URL

Botanical Name Portal

<http://www.anbg.gov.au/ibis/speciesLinks.html>

IPNI

<http://www.ipni.org/index.html>

Plants Database

<http://plants.usda.gov>

A.17 Transcriptome Related

Name and URL

Arizona Genomics Computational Lab

<http://www.agcol.arizona.edu/pave/cotton/>

Arizona Genomics Institute

<http://www.agcol.arizona.edu/cgi-bin/pave/Cotton/index.cgi>

Analysis of Early Events in Cotton Fiber Development

<http://www.cottongenomics.org>

Comparative Evolutionary Genomics of Cotton

<http://cottonrevolution.info>

Cotton Functional Genomics Center

<http://omics.hpcc.ttu.edu>

DFCI Cotton Gene Index

<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cotton>

NCBI (dbEST)

<http://www.ncbi.nlm.nih.gov/dbEST/>

NCBI (GEO)

<http://www.ncbi.nlm.nih.gov/geo/>

NCBI (UniGene)

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>

The Wendel lab

<http://www.eeob.iastate.edu/faculty/WendelJ/home.htm>

A.18 Biochemical Pathways

Name and URL
KEGG
http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=egra

A.20 Project and resource URLs mentioned in the text

Name and URL
MOBY Central
http://www.biomoby.org
CMap
http://gmod.sourceforge.net/cmap/index.shtml
Wikipedia: Web service introduction
http://en.wikipedia.org/wiki/Web_service
CEGC: genetically mapped objects portal
http://cotton.agtec.uga.edu/objects_portal/portal.aspx
CEGC: microarray information and data
http://cottonrevolution.info/microarray
Affymetrix: cotton array page
http://www.affymetrix.com/products_services/arrays/specific/cotton.affx
CEGC: microarray probes search tool
http://cotton.agtec.uga.edu/ProbePortal/portal.aspx
Vmatch
http://www.vmatch.de
W3C consortium
http://www.w3.org
W3C WSDL definition
http://www.w3.org/TR/wsdl
UDDI
http://www.uddi.org
NCBI: eUtils Web service
http://www.ncbi.nlm.nih.gov/entrez/query/static/esoap_help.html
Sequence ontology (SO)
http://www.sequenceontology.org/gff3.shtml
JGI: Poplar home
http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html

References

- Arpat, A. B., Waugh, M., Sullivan, J. P., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R. A. and Wilkins, T. A. (2004) Functional genomics of cell elongation in developing cotton fibers. *Plant Mol Biol.* 54:6, 911–929.
- Ball, C. A., Awad, I. A. B., Demeter, J., Gollub, J., Hebert, J. M., Hernandez-Boussard, T., Jin, H., Matese, J. C., Nitzberg, M., Wymore, F., Zachariah, Z. K., Brown, P. O. and Sherlock, G. (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Research.* 33, D580–D582.
- Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W-C, Ledoux, P., Rudnev, D., Lash, A. E., Fujibuchi, W. and Edgar, R. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Research.* 33, D562–D566.

- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. (2006) GenBank. *Nucleic Acids Research*. 34, D16–D20.
- Blenda, A., Scheffler, J., Scheffler, B., Palmer, M., Lacape, J-M, Yu J. Z., Jesudurai, C., Jung, S., Muthukumar, S., Yellambalase, P., Ficklin, S., Staton, M., Eshelman, R., Ulloa, M., Saha, S., Burr, B., Liu, S., Zhang, T., Fang, D., Pepper, A., Kumpatla, S., Jacobs, J., Tomkins, J., Cantrell, R. and Main, D. (2006) CMD: A Cotton Microsatellite Database Resource for *Gossypium* Genomics. *BMC Genomics*. 7:132.
- Buccella, A., Cechich, A. and Brisaboa, N. R. (2003) An Ontology Approach to Data Integration. *JCS&T*. 3:2.
- Carrere, S. and Gouzy, J. (2006) REMORA: a pilot in the ocean of BioMoby web-services. *Bioinformatics*. 22: 900–901.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*. 31:3497–3500.
- Dowell, R., Jokerst, R., Day, A., Eddy, S., Stein, L. (2001) The Distributed Annotation System. *BMC Bioinformatics* 2:7.
- Edgar, R. and Barrett, T. (2006) NCBI GEO standards and services for microarray data. *Nature Biotechnology*. 24, 1471 – 1472.
- Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005) The Sequence Ontology: A tool for the unification of genome annotations. *Genome Biol*. 6:R44.
- Estill, J. C., Bowers, J. E., Marler, B. S. and Paterson, A. H. (2003) BACMAN : BAC data management for high throughput physical mapping of genomes. *Plant, Animal & Microbe Genomes*. XI, 300.
- Frelichowski, J. E. Jr, Palmer, M. B., Main, D., Tomkins, J. P., Cantrell, R. G., Stelly, D. M., Yu, J. Z., Kohel, R. J. and Ulloa, M. (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. *Mol Genet Genom*. 275:479–491.
- Gingle, A. R., Yang, H., Chee, P. W., May, O. L., Rong, J., Bowman, D. T., Lubbers, E. L., Day, J. L. and Paterson, A. H. (2006) An Integrated Web Resource for Cotton. *Crop Sci*. 46, 1998–2007.
- Han, Z. G., Wang, C., Song, X. L., Guo, W. Z., Gou, J., Li, C., Chen, X. and Zhang, T. Z. (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor Appl Genet*. 112:430–439.
- Hsing, M. and Cherkasov, A. (2006) Integration of Biological Data with Semantic Networks. *Current Bioinformatics*. 1:3.
- Kawas, E., Senger, M. and Wilkinson, M. D. (2006) BioMoby extensions to the Taverna workflow management and enactment software. *BMC Bioinformatics*. 7:523.
- Kawashima, S., Katayama, T., Sato, Y. and Kanehisa, M. (2003) KEGG API: A Web Service Using SOAP/WSDL to Access the KEGG System. *Genome Informatics*. 14, 673–674.
- Lee, J., Hassan, O., Gao, W., Wei, N., Kohel, R., Chen, X. Y., Payton, P., Sze, S.H., Stelly, D., Chen, Z. (2006) Developmental and gene expression analyses of a cotton naked seed mutant. *Planta*. 223:3, 418–432.
- Lewis, S. E. (2004) Gene Ontology: looking backwards and forwards. *Genome Biology*. 6:103.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 27:1, 29–34.
- Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M. R. and Wipat, A. (2004). Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics*. 20, 3045–3054.
- Rong, J., Colette, A., Bowers, J. E., Brubaker, C. L., Chang, C., Chee, P. W., Delmonte, T. A., Ding, X., Garza, J. J., Marler, B. S., Park, C., Pierce, G. J., Rainey, K. M., Rastogi, V. K., Schulze, S. R., Trolinder, N. L., Wendel, J. F., Wilkins, T. A., Williams-Coplin, T. D., Wing, R. A., Wright, R. J., Zhao, X., Zhu, L., and Paterson, A. H. (2004). A 3347-locus

- genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*. 166:389–417.
- Rong, J., Feltus, A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., Paterson, A. H. Meta-analysis of Polyploid Cotton QTLs Shows Unequal Contributions of Subgenomes to a Complex Network of Genes and Gene Clusters Implicated in Lint Fiber Development. *Genetics*. *in press*.
- Ruiz, M., Rouard, M., Raboin, L. M., Lartaud, M., Lagoda, P. and Courtois, B. (2004) TropGENE-DB, a multi-tropical crop information system. *Nucleic Acids Research*. 1:32, D364–D367.
- Schneider, M., Bairoch, A., Wu, C. H. and Apweiler, R. (2005) Plant Protein Annotation in the UniProt Knowledgebase. *Plant Physiology*. 138, 59–66.
- Shah, N., Couronne, O., Pennacchio, L. A., Brudno, M., Batzoglu, S., Bethel, E. W., Rubin, E. M., Hamann, B. and Dubchak, I. (2004) Phylo-VISTA: interactive visualization of multiple DNA sequence alignments. *Bioinformatics*, 20, 636–43.
- Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. and Lewis, S. (2002) The Generic Genome Browser: A Building Block for a Model Organism System Database. *Genome Research*. 12, 1599–1610.
- Stevens, R.D., Robinson, A. J. and Goble, C. A. (2003) myGrid: personalised bioinformatics on the information grid. *Bioinformatics*, 19, i302–i304.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J. and Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 4: 41.
- The Gene Ontology Consortium. (2000) Gene Ontology: Tool for the unification of biology. *Nat. Genet.* 25: 25–29.
- The Gene Ontology Consortium. (2004) Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32, D258–D261.
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A. B., Sickler, B. A., Wilkins, T. A., Guo, J. Y., Chen, X. Y., Scheffler, J., Taliere, E., Turley, R., McFadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S. R., Pierce, M. L., Essenberg, M., Kim, H., Llewellyn, D. J., Dennis, E. S., Kudrna, D., Wing, R., Paterson, A. H., Soderlund, C. and Wendel, J. F. (2006) EST assembly for cotton genomics. *Genome Research*. 16(3):441–450.
- Udall, J. A., Cheung, F., Woodward, A. W., Swanson, J. M., Flagel, L., Hovav, R., Rapp, R. A., Nettleton, D., Lee, J. J., Gingle, A. R., Town, C., Chen, Z. J. and Wendel, J. F. (2007) Spotted cotton oligonucleotide microarrays for gene expression analysis *BMC Genomics*. 8:81.
- Whetzel, P.L., Parkinson, H., Causton, H.C., Fan, L., Fostel, J., Fragoso, G., Game, L., Heiskanen, M., Morrison, N., Rocca-Serra, P., Sansone, S-A, Taylor, C., White, J. and Stoeckert, Jr, C. J. (2006). The MGED ontology: a resource for semantics based description of microarray experiments. *Bioinformatics*. 22, 866–873.
- Wilkinson, M. D. and Links, M. (2002) BioMOBY: an open source biological web services proposal. *Brief Bioinform.* 3, 331–341.
- Yang, H., Wang, H. and Gingle, A.R. (2005) IntegratedMap: a Web interface for integrating genetic map data. *Bioinformatics*. 21, 2126–2127.
- Yang, H. and Gingle, A. R. (2005) OxfordGrid: a Web interface for pairwise comparative map views. *Bioinformatics*. 21:4307–4308.
- Yang, Y. H. and Paquet, A. C. (2005) Preprocessing Two-Color Spotted Arrays. In: R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry and S. Dudoit (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, pp. 49–69.

Bridging Classical and Molecular Cytogenetics of *Gossypium*

N'Guessan Olivier Konan, Jean-Pierre Baudoin, Angélique D'Hont,
and Guy Mergeai

Abstract Cotton (*Gossypium* spp.) is the leading natural fiber crop in the world. The genus *Gossypium* comprises a broad genetic base that has been and continues to be actively studied by cotton genetists and breeders. Cytogenetic tools play an important role in cotton genome research and cotton breeding. Based essentially on the observation of chromosome morphologies and the analysis of chromosome pairing, classical cytogenetics has contributed greatly to understanding cotton history, taxonomy and phylogeny, and has been a great help in cotton breeding programs designed to transfer desired genes from alien species into cultivated varieties. With the advent of molecular cytogenetics in the 1980s, the field of cytogenetics has been revolutionized. Beside an increase in the speed, sensitivity and specificity of conventional cytogenetic techniques, molecular cytogenetics offers opportunities to perform a variety of tasks not achievable by classical methods. These tasks include analysis of the distribution of repeated sequences along the genome, assignment of repetitive and single copy DNA sequences to positions on chromosomes, determination of the relationship between specific chromosomes and linkage groups, determination of the relationships between physical and genetic distances, differentiation of the genomes involved in hybrids, detection of alien DNA in introgressed lines, and others. We summarize the achievements of classical and molecular cytogenetic investigations in *Gossypium* and underline the relevance of bridging these approaches in *Gossypium* genetic studies and exploitation.

1 Introduction

Cytogenetics is a branch of genetics concerning every aspect of chromosomes that can be observed at the microscopic level. The first cytogenetic investigation in *Gossypium* started during the second half of the 19th century but it is since the

N.O. Konan (✉)

Gembloux Agricultural University (FUSAGx), Unité de Phytotechnie tropicale et d'Horticulture, 2 passage des Déportés, B-5030 Gembloux, Belgium
e-mail: konan.n@fsagx.ac.be; nguessanolivier@yahoo.fr

1920s (Denham 1924) that this science has become a powerful tool. Cytogenetic studies of cotton were first limited to examining chromosomes under an optical microscope, with staining techniques that enhance identification. Classical cytogenetics essentially allowed karyotyping, chromosome counting, detection of certain chromosomal abnormalities and analysis of chromosome pairing for assessment of genomic affinities. This conventional cytogenetics showed the existence of the diploid and polyploid species in *Gossypium* and greatly participated in the determination of the different cotton genome groups. Although this technique became very useful for *Gossypium* genome studies and cotton breeding, it was inadequate for detection of specific chromosomes or cryptic rearrangements. The remarkable development of molecular biology in the 1980s revolutionized the field of cytogenetics, with the advent of new methods which overcome many limitations of classical ones. This modern approach allows the precise physical localization of genes or DNA sequences on cytological preparations, enabling enormous progress in genome studies and breeding. Chromosome painting, study of chromosome exchanges and gene rearrangements, analysis of genome structure, development of cytomolecular maps, and monitoring of the transfer of agronomic traits are some of the interesting applications enabled by molecular cytogenetics. In spite of its success, modern techniques have not superseded classical cytogenetics; but have become an important complement that has bridged the gap between conventional cytogenetics and molecular genetic studies.

In this paper we examine the achievements obtained using classical and molecular cytogenetics to analyze and exploit *Gossypium*, emphasizing the interest of bridging these two complementary approaches.

2 Achievements of Classical Cytogenetics in Cotton

2.1 Principles of Classical Cytogenetics

Classical cytogenetics is a discipline of genetic that studies mitotic (from root tips or young leaves) or meiotic (from flower buds) chromosomes under an optical microscope using staining or banding techniques for their identification. Generally, metaphase chromosomes are studied because they are the most condensed and the most visible. The stains usually used to visualize chromosomes are ferric haematoxylin, crystal violet, acetic orceine, schiff reagent, and carmine acetic acid; they allow uniform staining of chromosomes. Classical staining techniques are usually used for chromosome counting, studies of chromosome morphology, detection of chromosomal modification and analysis of chromosome pairing. In banding, chromosomes are stained with Giemsa after various treatments (enzymatic, denaturing agent, heat...) that allow staining of chromosomes with a succession of light and dark bands specific to each chromosomal pairs. The most important banding techniques are

G (Giemsa), R (reverse), BrdU (5-Bromodéoxyuridine), C (centromeric) and T (telomeric) banding. Banding techniques are useful for identifying whole chromosomes accurately (chromosomes present in interspecific hybrids, addition or substitution lines, aneuploid stocks) and for detecting chromosomal rearrangements; but this technique is laborious to apply and was much more rarely used in cotton classical cytogenetic investigations than the staining techniques.

2.2 Contributions of Classical Cytogenetics to Taxonomic and Phylogenetic Studies of *Gossypium*

2.2.1 Classical Cytogenetics in the Identification of *Gossypium* Genomes and the Understanding of *Gossypium* Phylogeny

Classical cytogenetic studies greatly contributed to current knowledge of *Gossypium* phylogeny. In the late 1920s and early 1930s, classical cytogenetic analyses of species and genome types of *Gossypium* were performed by a number of investigators (Endrizzi, Turcotte and Kohel 1985). It was Nikolajeva (1923) and Denham (1924), who the first demonstrated that there were exactly two chromosome numbers (26 and 52 chromosomes) in the genus *Gossypium*. Based on this observation, Denham (1924) divided the species of *Gossypium* into diploids ($2n=26$) and tetraploids ($2n=52$). The cytogenetic studies of Denham represent a decisive step in cotton systematics because they opened the door to a long series of works that contributed greatly to the classification of the genus *Gossypium*. Several subsequent workers reported chromosome numbers of 26 and 52 for many additional *Gossypium* species including wild and cultivated forms occurring in America, Asia, Africa, and Australia (Youngman and Pande 1927; Harland 1928; Banerji 1929; Baranov 1930; Kearney 1930; Longley 1933; Skovsted 1933, 1934a,b, 1935b; Webber 1934a,b, 1935, 1939; Wouters 1948). Currently the genus *Gossypium* includes 45 diploid and 5 tetraploid species (Wendel and Cronn 2003).

Meiotic studies of the diploid species and their hybrids allowed classification of cotton species into different genome groups. Beasley (1940, 1942), based on chromosomes pairing at meiotic metaphase I in F1 interspecific hybrids, was the first to propose a genomic classification of the diploid species. Some species exhibited highly regular pairing of chromosomes in their hybrids, whereas other species showed highly irregular chromosome pairing. Beasley used the frequency of paired chromosomes to measure the affinities between species and to distinguish the different genomes. He defined five genome groups designed by the capital letters A, B, C, D, and E. In 1950, Brown and Menzel perfected the system of symbols proposed by Beasley. Similar genomes are designated by the same capital letter and closely related genomes are distinguished by a numerical subscript after each letter of that class. Eight diploid and one tetraploid genome group(s) are now recognized in *Gossypium* (Table 1). The A to E genomes were assigned by Beasley (1940), F and G genomes were assigned

Table 1 Localization and species composition of the *Gossypium* genome groups

Genomes	Localization	Species
A	Africa and Asia	<i>G. herbaceum</i> (A ₁), <i>G. arboreum</i> (A ₂)
B	Africa	<i>G. anomalum</i> (B ₁), <i>G. triphyllum</i> (B ₂), <i>G. barbosanum</i> (B ₃), <i>G. capitis-viridis</i> (B ₄)
C	Australia	<i>G. sturtianum</i> (C ₁), <i>G. robinsonii</i> (C ₂), <i>G. nandewarense</i>
D	America	<i>G. thurberi</i> (D ₁), <i>G. armourianum</i> (D ₂₋₁), <i>G. harknessii</i> (D ₂₋₂), <i>G. davidsonii</i> (D _{3-d}), <i>G. klotzschianum</i> (D _{3-k}), <i>G. aridum</i> (D ₄), <i>G. raimondii</i> (D ₅), <i>G. gossypoides</i> (D ₆), <i>G. lobatum</i> (D ₇), <i>G. trilobum</i> (D ₈), <i>G. laxum</i> (D ₉), <i>G. turneri</i> (D ₁₀), <i>G. schwendimanii</i> (D ₁₁)
E	Arabia and Africa	<i>G. stocksii</i> (E ₁), <i>G. somalense</i> (E ₂), <i>G. areysianum</i> (E ₃), <i>G. incanum</i> (E ₄), <i>G. benadirensis</i> , <i>G. bricchettii</i> , <i>G. vollesenii</i>
F	Africa	<i>G. longicalyx</i> (F ₁)
G	Australia	<i>G. bickii</i> (G ₁), <i>G. australe</i> , <i>G. nelsonii</i>
K	NW Australia	<i>G. costulatum</i> , <i>G. cunninghamii</i> , <i>G. enthyale</i> , <i>G. exgium</i> , <i>G. nobile</i> , <i>G. pilosum</i> , <i>G. populifolium</i> , <i>G. pulchellum</i> , <i>G. rotundifolium</i> , <i>G. sp.novum</i>
AD	America	<i>G. hirsutum</i> (AD) ₁ , <i>G. barbadense</i> (AD) ₂ , <i>G. tomentosum</i> (AD) ₃ , <i>G. mustelinum</i> (AD) ₄ , <i>G. darwinii</i> (AD) ₅

The lack of numerical subscript after the letter designating the genome of a species indicates that detailed cytogenetic studies have not been carried out yet for it and that its genomic classification is provisional.

respectively by Phillips and Strickland (1966) and Edwards and Mirza (1979). Table 1 shows the different genome groups in *Gossypium*, their location and the species they contain.

Moreover, classical cytogenetic studies revealed cotton chromosomes to vary widely in size (Stephens 1947, Katterman and Ergle 1970): i) the C genome has very large chromosomes; ii) E and F genomes have large chromosomes that are slightly larger than those of the A and B genomes; iii) the B genome has large chromosomes, some of which are slightly larger than those of the A genome; iv) the A genome has moderately large chromosomes; v) the G genome has moderately large chromosomes but smaller than those of the A genome; vi) the D genome has the smallest chromosomes. Cytogenetic studies showed that chromosome morphology is similar among closely related species, and related species form hybrids with normal meiotic pairing; whereas, hybrids from relatively distant species show meiotic abnormalities. It has been possible to assess the genome relationships and level of divergence based on the frequency of unpaired or univalent chromosomes in intra and intergenomic hybrids of the diploid species (Table 2). No cytogenetic studies were carried out so far on intergenomic diploid hybrids involving the F-genome species *G. longicalyx*. The relationships of this species with the other *Gossypium* genomes can however be inferred from the work of Philips and Strickland (1966) who compared the meiotic pairing in triploids and hexaploids resulting from crosses between the

Table 2 Average univalent frequency in intergenomic hybrids of diploid *Gossypium*

Intergenomic hybrid	Univalents per cell
A x B	2.82
A x C	8.50
A x D	13.98
A x E	17.13
A x G	16.00
B x C	11.17
B x D	18.19
B x E	22.35
C x D	13.10
C x E	24.68
D x E	25.15
D x F	21.60
G x C	3.84

Source : Endrizzi et al. (1985), N'Dungo et al. (1988b)

amphidiploids and diploid species belonging to all the different genomes (except K-genome). They put in evidence that the pairing affinities existing between the F-genome and the A_h subgenome chromosomes were higher than the ones observed between A_h and B.

Classical cytogenetics contributed to the understanding of the origin of the diploid species of cotton. Indeed, studies based on secondary associations that are visible during meiotic metaphase in diploid cotton (Davie 1933; Skovsted 1933; Abraham 1940; Brown and Menzel 1952), karyotype analysis by Edwards et al. (1979), and the application of BrdU-Hoechst-Giemsa chromosome banding techniques to diploid *Gossypium* (Muravenko, Fedotov, Punina, Fedorova, Grif and Zelenin 1998) suggested a paleopolyploid origin of the diploid *Gossypium*. The diploid species may have come from a single ancestral taxon, that was polyploid with the juxtaposition of 6 + 7 chromosomes.

2.2.2 Classical Cytogenetics in the Study of Tetraploid *Gossypium*

Classical cytogenetics helped to answer some questions raised by the discovery of tetraploid cottons. At first, classical cytogenetic studies demonstrated that the tetraploid cottons are true allotetraploids that contain two resident genomes, an A-genome from Africa or Asia, and a D-genome similar to those found in the American diploids. Beal (1928) had noted that the chromosomes of the allotetraploid species varied in size at metaphase I, and Davie (1933) had recorded a variation in somatic chromosome length of *G. hirsutum* ranging between 0.8 and 1.8 μm . Skovsted (1934a, b, 1935a) made the important observation that the tetraploid cottons had 13 large chromosomes with a

mean length of 2.26 to 2.36 μm and 13 small chromosomes with a mean length of 1.25–1.45 μm . Based on the meiotic behavior of a number of interspecific hybrids, and on the comparative size of the chromosomes of species in the genus, Skovsted concluded the tetraploid cottons were amphidiploids that originated by doubling of the nonhomologous chromosomes of two species with $n = 13$, one of which was similar to the large A chromosomes and the other of which was similar to the small D chromosomes (Endrizzi 1985). Webber (1939) and Beasley (1940, 1942) confirmed the hypothesis provided by Skovsted (1934b). Thus, by the early 1940s, it was firmly established, by classical cytogenetic studies, that the tetraploid cottons were allotetraploids that originated from combining A and D genomes.

The main contributions of classical cytogenetic studies to the understanding of *Gossypium* evolution are presented in Fig. 1.

2.2.3 Classical Cytogenetics in the Study of Translocations in *Gossypium*

Meiotic studies of cotton hybrids by classical cytogenetics revealed the presence of multivalent chromosome configurations at metaphase I. Beasley (1942) referred to the multivalents as translocations. Brown et al. (1950), Gerstel (1953), and Menzel and Brown (1954) observed also that the A genome of diploid species and the A_n subgenome of the allotetraploid differ from each other by chromosomal interchanges. The discovery of translocation in

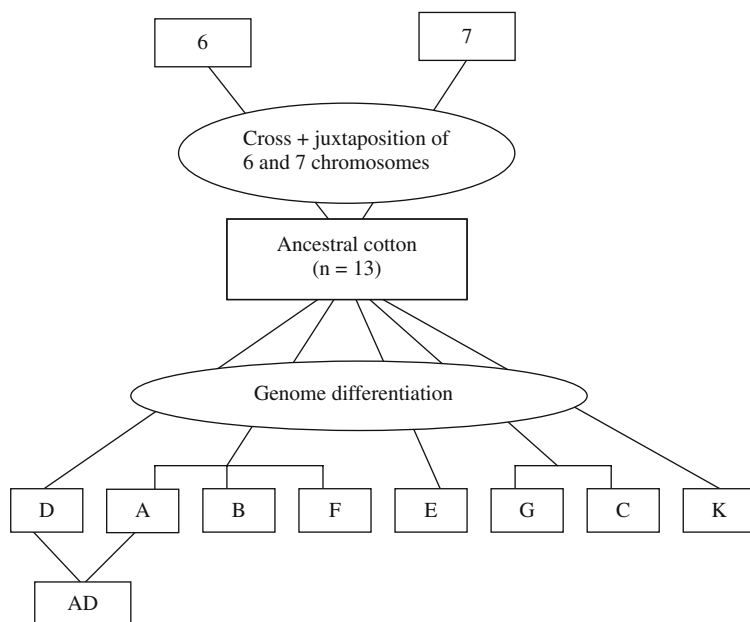


Fig. 1 Evolutionary history of *Gossypium* deduced from classical cytogenetic studies

Gossypium thanks to classical cytogenetic studies, allowed the identification of individual chromosomes. Consequently many translocations were induced in cotton by irradiation (Brown 1950; Menzel et al. 1954) and a set of 62 homologous translocation lines was developed (Brown 1980). Based on these translocation lines, 25 of the 26 chromosomes of *G. hirsutum* have been distinguished and numbered (Brown 1980). The A subgenome chromosomes were numbered A1-A13 (H1-H13) and the D subgenome chromosomes D1-D13 (H14-H26). Chromosome 26 has not been involved in a known translocation and it was identified through a process of elimination (Endrizzi et al. 1985). The translocation stocks constitute the only complete set of cytogenetic markers for the *G. hirsutum* genome (Wang, Song, Han, Guo, Yu, Sun, Pan, Kohel and Zhang 2006). The cytological procedure for assigning chromosomes in translocations to their respective genomes was described by Menzel (1955) and Endrizzi et al (1985). The identity of the chromosomes in each of the 62 translocations was determined in most cases by intercrossing the homozygous translocations and examining chromosomes pairing to determine whether the same or different chromosomes were involved (Brown 1980; Endrizzi et al. 1985). In some cases, stocks that were monosomic for known chromosomes were used to identify chromosomes in the translocations and vice versa (Endrizzi et al. 1985).

The translocation breakpoints with known locations can serve as reference markers in positioning genes and linkage groups on their chromosome (Menzel, Richmond and Dougherty 1985). Menzel and Brown (1978a,b) and Brown, Menzel, Hasenkampf and Naqi (1981) determined the arm location of breakpoints in the translocated chromosomes of many translocations. Menzel et al. (1985) assigned 115 translocation breakpoints to their chromosome arms in 58 translocations that involve exchanges between only two chromosomes. Using chiasma frequencies in specific chromosome regions, they estimated the recombination map length of each chromosome and the distance of each breakpoint from its centromere. Using these data they constructed the first cotton genome map with reference points on all the chromosomes, except chromosome 26. Thus, using classical cytogenetics, it has been possible to construct a recombination map of the cotton genome based on chiasma frequencies in chromosome regions defined by the breakpoints in 58 reciprocal chromosome translocations.

2.3 Classical Cytogenetics in Cotton Breeding

Once taxonomic and phyletic relationship among *Gossypium* species became better understood, classical cytogenetic studies were mostly used in breeding programs. All cotton genomes are important genetic resources that can contribute valuable genes for fibre quality, resistance to diseases and insect pests, tolerance to abiotic stress, and cytoplasmic genes conditioning male sterility along with nuclear restorer genes (Ndungo, Demol and Maréchal 1988c; Stewart 1995). In interspecific breeding programmes, classical cytogenetics is

an important discipline because it allows i) the understanding of genomic affinities in order to facilitate the planning of effective hybridization programs designed to transfer desired genes from alien species into cultivars; ii) the rational exploitation of intraspecific, interspecific, or induced variability; and iii) the explanation and resolution of some problems met by breeders such as instability or sterility.

In cotton breeding, classical cytogenetic studies focus mainly on meiotic analysis. Observations concern mainly the chromosomal configurations present at metaphase I. The goal is to establish a mean karyological formula from the analysis of a representative sample of pollen mother cells. The operation consists of the listing, for each mother cell, of the number of the different types of configurations observed (univalents, bivalents, trivalents, quadrivalents, etc.) and the total number of chiasmata. The average of each configuration is calculated. The deviation compared to the normal meiotic behavior (Fig. 2) can reveal the stability level of a hybrid, the type and the level of homology between the parental species (allowing the assessment of the possibilities of recombination) and certain structural rearrangements (deletions, duplications, translocations). A precise monitoring of alien chromosome segment introgression in interspecific hybrid progenies cannot however be obtained with classical cytogenetic methods.

A synthesis of the meiotic chromosome configurations observed in trispecific allotetraploids, bispecific triploids, and bispecific allohexaploids is presented in Tables 3, 4 and 5, respectively. The main lessons that can be drawn from the application of cytogenetic investigations in the framework of cotton interspecific breeding programs are summarized below.

The immediate expression of almost complete homoeologous pairing in the raw amphidiploid hybrids obtained by chromosome doubling of diploid or triploid bispecific hybrids and the lack of prevention of synapsis of homoeologous chromosomes during prophase in cotton haploids (Endrizzi et al. 1985)

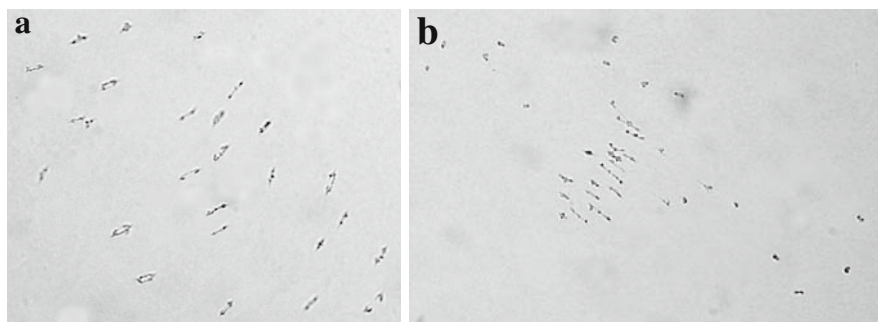


Fig. 2 Meiotic configuration at metaphase I showing regular pairing with 26 bivalent in *G. hirsutum* (a) and irregular pairing with univalent and multivalent in a trispecific hybrid $[(G. hirsutum \times G. thurberi)^2 \times G. longicalyx]$ (b)

Table 3 Meiotic chromosome configurations in trispecific hybrids of *Gossypium* obtained with classical cytogenetic analysis

Hybrid combinations	Chr. No.	Chromosome configuration								References
		I	II	III	IV	V	VI	VIII	X	
[(<i>G. arboreum</i> x <i>G. thurberi</i>) ² x <i>G. hirsutum</i>]	52	0.88	20.93	0.62	1.01	0.14	0.27	0.02	0.02	Brown and Menzel (1950)
[(<i>G. hirsutum</i> x <i>G. arboreum</i>) ² x <i>G. harknessii</i>]	52	1.37	22.36	0.31	1.09	0.01	0.08	0.00	0.00	Brown et al. (1950)
[(<i>G. hirsutum</i> x <i>G. herbaceum</i>) ² x <i>G. harknessii</i>]	52	1.84	22.14	1.14	0.54	0.00	0.05	0.00	0.00	Brown et al. (1950)
[(<i>G. thurberi</i> x <i>G. anomalum</i>) ² x <i>G. hirsutum</i>]	52	11.45	17.13	1.22	0.50	0.04	0.03	0.00	0.00	Louant & Maréchal (1975)
[(<i>G. hirsutum</i> x <i>G. anomalum</i>) ² x <i>G. harknessii</i>]	52	17.72	16.24	0.46	0.10	0.00	0.00	0.00	0.00	Louant et al. (1975)
[(<i>G. hirsutum</i> x <i>G. raimondii</i>) ² x <i>G. sturtianum</i>]	52	13.64	17.04	0.85	0.35	0.00	0.07	0.00	0.00	Vroh bi, Hau, Baudoin and Mergeai (1999)
[(<i>G. thurberi</i> x <i>G. sturtianum</i>) ² x <i>G. hirsutum</i>]	52	14.55	15.68	0.91	0.35	0.00	0.25	0.00	0.00	Vroh bi et al. (1999)
[(<i>G. arboreum</i> x <i>G. bickii</i>) ² x <i>G. hirsutum</i>]	52	41.04	4.54	0.57	0.04	0.00	0.00	0.00	0.00	Shuijin & Biling (1993)
[(<i>G. hirsutum</i> x <i>G. thurberi</i>) ² x <i>G. longicalyx</i>]	52	14.13	15.10	1.03	0.9	0.03	0.13	0.00	0.00	Konan, D'Hont, Baudoin and Mergeai (2007)

Table 4 Mean meiotic chromosome configurations in bispecific triploid hybrids of *Gossypium* obtained with classical cytogenetic analysis

Triploid	Chr. No.	Chromosome configuration						References
		I	II	III	IV	V	VI	
AD x A	39	13.00	11.33	0.00	0.33	0.00	0.33	Gerstel (1953); Baranov (1930); Skovsted (1934a)
AD x B	39	25.05	6.34	0.33	0.06	0.00	0.00	Poisson (1970)
AD x C	39	26.32	5.41	0.55	0.04	0.00	0.00	Skovsted (1937); Maréchal (1974)
AD x D	39	14.36	11.72	0.37	0.03	0.00	0.00	Skovsted (1937); Iyengar (1944); Endrizzi (1957); Boza and Madoo (1941); Kammacher (1960) Menzel and Brown (1954)

Table 5 Mean meiotic chromosome configurations in bispecific hexaploid hybrids of *Gossypium* obtained with classical cytogenetic analysis

Hexaploid	Chr. No.	Chromosome configuration								References
		I	II	III	IV	V	VI	VII	VIII	
AD x A	78	2.13	34.63	1.38	0.63	0.00	0.00	0.00	0.00	Iyengar (1944)
AD x B	78	2.19	36.17	0.38	0.50	0.02	0.03	0.00	0.01	Poisson (1970); Iyengar (1944), Brown and Menzel (1952);
AD x C	78	1.68	36.54	0.32	0.57	0.00	0.00	0.00	0.00	Brown and Menzel (1952)
AD x D	78	1.47	30.07	0.82	3.38	0.02	0.03	0.00	0.00	Brown and Menzel (1952); Iyengar (1944);
AD x E	78	1.26	38.07	0.09	0.08	0.00	0.00	0.00	0.00	Brown and Menzel (1952); Maréchal (1972); Schwendiman, Koto and Hau (1980)
AD x F	78	1.47	35.28	0.28	1.22	0.03	0.02	0.00	0.00	Phillips and Strickland (1966); Schwendiman et al. (1980)

suggest that if a gene functioning like *Ph1* in wheat for regulating bivalent pairing (Riley and Chapman 1958; Sears and Okamoto 1958; Riley, Chapman and Kimber 1960) probably exist in the allotetraploid cotton species, in accordance with the postulate made by Kimber (1961), its impact on the prevention of intergenomic pairing is much lower in these species than in wheat. Consequently, the introgression of alien chromosome segments in upland cotton in the framework of interspecific breeding programs should be easier in *Gossypium* than in *Triticum*.

All trispecific hybrids are obtained through the development of an intermediary hybrid, which can be either allotetraploid or allohexaploid. In these crossing schemes, besides the choice of the bridge species, the breeder cannot do much to control the level and the nature of genetic material exchanges between the different genomes that are combined in the tri-species structure. Cytogenetic analyses carried out in programs involving tri-species hybrids reveal the importance of choosing a diploid species belonging to genome D as a bridge species when creating such materials. Indeed, because chromosomes prefer auto-syndetic pairing at meiosis, recombination is low for chromosomes with low homology. Genome D chromosomes being much smaller than the chromosomes of the other diploid genomes, their pairing affinities with the latter are very low. It means that if a species other than a D-genome species is used as bridge to create the trispecific hybrid, most of the chromosomes of subgenome D_h from *G. hirsutum* will remain unassociated at metaphase I and it will be almost impossible to obtain fertile progeny by backcrossing the trispecific hybrid to *G. hirsutum*. This is illustrated by the Metaphase I chromosome configurations observed by Shuijing and Biling (1993) in the ABH [$(G. arboreum \times G. bickii)^2 \times G. hirsutum$, A_hA₂D_hG₁] hybrid and what was observed in the TSH [$(G. thurberi \times G. sturtianum)^2 \times G. hirsutum$, A_hC₁D_hD₅] and HRS [$(G. hirsutum \times G. raimondii)^2 \times G. sturtianum$, A_hC₁D_hD₅] trispecific hybrids (Mergeai, Baudoin and Vroh Bi. 1997). A rather high pairing frequency was observed in TSH and HRS hybrids (Table 3) which gave rise to fertile progeny while the number of bivalents and multivalents was very low in the ABH hybrid ($2n=4x=52=41.01 \text{ I} + 4.54 \text{ II} + 0.57 \text{ III} + 0.41 \text{ IV}$) from which no viable seeds could be produced. The application of growth regulators (50 mg.l⁻¹ naphthoxy-acetic acid + 100 mg.l⁻¹ gibberellic acid) to avoid capsule shedding after pollination and the *in vitro* rescue of mature embryos (Vroh Bi, Baudoin, Hau, and Mergeai 1999) allowed the exploitation of tri-species hybrids in which the donor species did not belong to the diploid genomes that are genetically close to the A_h and D_h subgenomes of *G. hirsutum*. This is notably the case for *G. sturtianum* (Genome C), *G. longicalyx* (Genome F) and *G. areysianum* (Genome E), which were included in fertile trispecific hybrids. Due to unfavorable linkages that exist between agronomic and fiber quality traits in the trispecific hybrids involving *G. thurberi* as bridge species (Demol 1966; Miller and Rawlings 1967; Meredith and Bridge 1971), it is recommended to use *G. raimondii* for this purpose.

The tri-specific pathway using a D-Genome species as bridge is interesting because in such allotetraploid combinations the A_h chromosomes have no auto-syndetic partners and theoretically should pair with the chromosomes of the donor species. However, the successful use of trispecific synthetic tetraploids requires generally a large effort to produce fertile progeny and to eliminate the undesirable genetic material contributed by the diploid donor and bridge species (Vroh Bi et al. 1999). Although the frequency of homologous recombination between the donor species chromosomes and the A_h - or D_h -Genome chromosomes may be lower in bi-specific than tri-specific derivatives, the bi-specific pathway theoretically offers the possibility of generating more progeny in the same amount of time and thus to capture more homologous recombination events. Moreover, in direct exploitation of bi-specific hybrids through backcrossing the hexaploids to *G. hirsutum*, recombinant chromosomes are far more likely to be incorporated into fertile plants. This last method has also the advantage of allowing some control of the intensity of genetic exchanges in interspecific hybrids at the hexaploid and monosomic addition stages (Louant, Maréchal and Baudoin 1977). In such fertile interspecific structures, it is possible to accumulate recombination events over generations. Genetic material transfer from the wild diploid species to the cultivated amphidiploid is enhanced through spontaneous production, during the successive hexaploid generations, of intergenomic exchanges, which improve the affinities between the genomes. For each interspecific combination, there should exist an ideal number of hexaploid generations that would allow the production of the optimal level of intergenomic exchanges adapted to the genetic nature of the diploid species relative to *G. hirsutum*. This trend increases the chance of capturing a trait in the subsequent backcrosses to *G. hirsutum* of the allohexaploids. The interchanges occurring during successive allohexaploid generations lead to the production of pentaploid types differing according to the extent of recombinations that occurred in their interspecific parent material. For the genomes that are closely related to the amphidiploid subgenomes (A, D), the number of allohexaploid generations is limited to one or two because the very high level of genetic recombinations that occur every generation leads quickly to a complete sterility of the hybrid. The optimal number of hexaploid generations for F, B, C and G-genome species still needs to be determined. It should be inversely proportional to the multivalent frequencies observed at metaphase I in each hybrid combination (Table 5). The multiplication of hexaploid generations in bispecific hybrids involving E-genome species is not useful because in such hybrids the very low intergenomic pairing frequencies occurring at the hexaploid stage remain unchanged over generations. Monosomic addition plants are other fertile structures on which the breeder can exert some control regarding the introgression into the tetraploid cotton genome of characters from the supernumerary chromosome brought by the diploid donor species. Beside monosomic and disomic addition plants, the selfing of a monosomic addition line gives rise to euploid materials which can be introgressed by chromosomal fragments of the donor species. The extent of this introgression

depends on pairing affinities existing between the alien supernumerary chromosome and the genome of the recipient species. Once an agronomic trait of interest is identified in a monosomic addition line, plants carrying the donor diploid species supernumerary chromosome of this line can be used to conduct chromosome specific introgression by selfing them until incorporation of the desired trait into *G. hirsutum* (Hau 1981).

In order to isolate a large number of monosomic addition plants when following the bispecific introgression pathway, it is recommended, provided its pollen fertility is sufficiently high, to use the pentaploid as male parent in the backcross to *G. hirsutum* (Ahoton, Lacape, Baudoin and Mergeai 2003). In this type of cross, the progeny obtained contain a rather high proportion of monosomic addition plants (about 10 %) plus a large majority of euploid materials and almost no other aneuploid genotypes. When one uses the pentaploid as female parent in the backcross to *G. hirsutum*, most of the progeny are auto-sterile plants carrying several alien chromosomes. This trend is observed in the progeny of various crosses carried out between *G. hirsutum* and pentaploid hybrids involving diploid species of C, G, E and F-genome. It indirectly confirms the better tolerance of female gametes for multiple alien chromosome addition in their nucleus, and the better competitiveness of male gametes carrying only one additional alien chromosome compared to those carrying several alien chromosomes. In the progeny of the pentaploid and the monosomic addition lines, each alien chromosome addition is characterized by a particular transmission rate, which is chromosome specific. These variations can be explained by differences between the alien addition chromosome and its homology with the *G. hirsutum* genome. It can also be explained by various factors acting on the viability of aneuploid male and female gametes, on aneuploid zygote development, on aneuploid seed germination, and on the survival of plants carrying an alien supernumerary chromosome.

3 Achievements of Molecular Cytogenetics in Cotton

3.1 Principles of Molecular Cytogenetics

The coupling of molecular technologies with cytogenetics gave rise to molecular cytogenetics, largely replacing chemical stains with molecular probes and *in situ* hybridization (ISH) techniques, notably FISH (fluorescent *in situ* hybridization). FISH is based on the property of a sequence of DNA (probe) to hybridize to a complementary DNA (target). The technique involves labelling molecular probes with fluorochromes and hybridization of these fluorescently labelled probes to unique DNA sequences *in situ*. Probe detection is accomplished by ultraviolet-light excitement of fluorochromes which are directly attached to probe DNA, producing fluorescent signals that are inspected using a filter-equipped epifluorescence microscope and computer software. The different

types of probes commonly used are gene-specific probes, repetitive sequence probes, DNA clone probes (bacterial artificial chromosomes, BAC-FISH) and total genomic DNA probes (genomic *in situ* hybridization, i.e. GISH). FISH probes can range in size from over 100 kb, to less than 1 kb.

Fluorescent *in situ* hybridization is a powerful tool for genome analysis and genetic manipulation. By this sensitive method, molecular cytogenetics allows the exploration of genetic material, rendering possible the physical visualization of genes, DNA sequences, or specific chromosomes under microscope on cytological preparations. Consequently, molecular cytogenetics is a means to perform a variety of tasks such as assigning repetitive and single copy DNA sequences to positions on chromosomes, assigning molecular marker linkage groups to specific chromosomes and chromosome arms, detecting alien chromosomes in hybrids, detecting alien DNA in introgressed lines, and detecting and numbering sites of transgene inserts.

3.2 Physical Mapping of Repeated Sequences in *Gossypium*

Repeated sequences could be used as chromosome markers, cytological landmarks having value in chromosome identification (Mukai, Friebe and Gill 1992; Rayburn and Gill 1986; Tsujimoto, Mukai, Akagawa, Nagaki, Fujigaki, Yamamoto and Sasakuma 1997). Repeated sequences can be divided into two types based on their organization and distribution pattern. Tandem arrays are localized as clusters along the chromosomes, and dispersed repeats are interspersed with unrelated repeats and low-copy DNA over much of the genome. Among the most thoroughly investigated repetitive DNA sequences in plant species are the ribosomal RNA genes (rDNA), including 18S-26S and 5S rDNA tandem repeats (Schmidt et al. 1994; Cabrera et al. 1995; Castilho and Heslop-Harrison 1995). The rRNA genes are organized in tandem arrays within the nucleolar organizer regions (NORs). The ribosomal RNA genes can be used as probes for physical mapping in higher plants because they are arranged in tandem arrays clustered at a few sites. Visualization of these rRNA genes by FISH can provide a number of chromosomal markers to elucidate chromosome evolution and species interrelationships, including evolution of polyploid species.

In cotton, Bergey, Stelly, Price, and McKnight (1989) detected enzymatically three major 18S-26S rDNA sites (NORs) in *G. hirsutum* by meiotic ISH. Price, Stelly, McKnight, Scheuring, Raska, Michaelson and Bergey (1990) used a biotin-labeled cloned fragment of 18S-28S ribosomal DNA from soybean to hybridize DNA in meiotic chromosomes of *G. hirsutum* by *in situ* hybridization. Analysis of *in situ* hybridization to metaphase I meiocytes from two translocation heterozygotes and monosomics involving chromosome 9 indicated that a cluster of ribosomal RNA cistrons is on chromosome arm 9L, arguably the first

molecular marker mapping in cotton. Three years later, Crane, Price, Stelly and Czeschin (1993) discovered other rDNA sites. In 1996, Hanson, Islam-Faridi, Percival, Crane, Ji, McKnight, Stelly and Price used high-sensitivity mitotic FISH on somatic chromosomes to reveal at least six additional 18S-26S rDNA loci, but none of these latter loci have been mapped. Ji, De Donato, Canel, Raska, Islam-Faridi, McKnight, Price and Stelly (1999) detected four more minor sites by meiotic FISH. It was Ji et al. (1999) who mapped and integrated all the known rDNA loci into the translocation breakpoint map of Menzel et al. (1985) by using meiotic FISH to quadrivalents of translocation heterozygotes. The detected 18S-26S rDNA loci were mapped to the right arms of chromosomes 8, 9, 15, 17, 19, 20, and 23 and the left arms of chromosomes 5, 11, 12, and 14. Using the rDNA loci as common reference points, Ji et al (1999) detected several erroneous arm assignments in the previously published map of heterozygous translocation breakpoints. Therefore, the use of meiotic FISH constitutes an important method for accurate physical mapping. The multiplicity of rDNA sites and consistent detection of many of them makes them useful in molecular cytogenetic studies of *G. hirsutum*. First, the three major NOR sites, four intermediate 18S-26S rDNA loci, and two 5S rDNA loci provide multiple, consistently detectable markers for karyological studies. Crane et al. (1993) indicated that the three major sites and the intermediate site on chromosome 7 could serve to define arm locations and thus facilitate translocation-based mapping of new loci in 18 of the 26 cotton chromosomes. Analogously, the other localized intermediate rDNA sites on chromosomes 5, 17, and 19 allow the definition of arm locations of new loci in all other chromosomes, except chromosomes 22 and 26. Chromosome 22 is associated with only one translocation, in which a minor 18S-26S site marks arm 20R. Chromosome 26 is not involved in any of the known translocations, so its arms are cytogenetically distinguishable only by monotelodisomy. The distribution of rDNA genes on chromosomal arms suggests that rRNA genes in short arms may be more likely to organize nucleoli. The map of rDNA locations will facilitate site-specific analysis of rRNA gene function and rDNA evolution, using methods for identification of rDNA-containing bacterial artificial chromosomes (Woo 1996).

In cotton, chromosome translocations are the primary sources for chromosome identification. In 1997, Ji, Raska, McKnight, Islam-Faridi, Crane, Zwick, Hanson, Price and Stelly used meiotic FISH to analyze a new monosome of *G. hirsutum*. Painting with A2-genome DNA revealed the monosome's D-subgenome origin. DAPI-PI staining showed that the monosome carries a major NOR, delimiting it to the major NOR-bearing chromosomes of the D-subgenome, i.e., 16 or 23. Dual-color FISH with 5S and 18S-28S rDNAs indicated that the monosome contains separate major clusters of each of these two tandemly repeated rDNA elements, thus delimiting the monosome to chromosome 23, for which the Cotton Cytogenetic Collection previously lacked any sort of deficiency.

3.3 Identification of Structural Abnormalities in *Gossypium*

Chromosome structural abnormalities, such as duplication-deficiencies (dp-dfs), can be used, as with other hemizygous aneuploids, to assign genes or molecular markers to their respective chromosomes. Well established translocation stocks exist in cotton (*G. hirsutum*). The cotton cytogenetic collection includes a set of 58 simple reciprocal translocations in upland cotton (*G. hirsutum*), which constitutes the primary resource for chromosome manipulation and identification. Breakpoints of those translocations collectively affect 25 of the 26 chromosomes, and most have been mapped relative to one another and to their respective centromeres and telomeres, based on interstitial and distal chiasma frequencies in translocation heterozygotes (Menzel and Richmond 1985; Menzel and Dougherty 1987; Menzel, Richmond and Dougherty 1987; Menzel et al. 1985). However, conventional means of dp-df identification are often difficult to apply. Ji, Raska, De Donato, Islam-Faridi, Price, and Stelly (1999) demonstrated that by means of meiotic fluorescent *in situ* hybridization, the identification of the four types of dp-dfs is greatly facilitated, proving that fluorescence *in situ* hybridization (FISH) overcomes the limitations of classical cytogenetics and can be used to identify the structural abnormalities. That is important since dp-dfs can be used to assign genes or molecular markers to their respective chromosomes, and thus can facilitate the integration of physical and recombination maps.

3.4 Physical Mapping of Agronomically Important Genes

In crops, the physical mapping of genes, especially those of high agronomic value, is a valuable step in research programs dealing with the improvement of cultivars through hybridization and gene manipulation. Physical and molecular cytogenetic maps of cultivated plants have great practical and research value. They open up the prospect of producing chromosome-specific DNA libraries from parts of chromosomes containing valuable genes. Another key aim is to integrate the physical and (or) cytological map of a crop plant with its genetic linkage map. Genes and markers are then assigned to linkage groups, each of which corresponds to a single chromosome (Jiang and Gill 1994).

Agronomically important genes are mostly unique or low copy sequences. The information about the exact physical location of agronomically important genes is useful in breeding programs as well as in understanding the organization of genomes. In plants, *in situ* hybridization techniques have been used mainly for mapping repetitive DNA sequences and multicopy gene families. Mapping of low or single-copy sequences has proven difficult in plants compared to humans. On human chromosomes, single copy sequences as small as 1 kb can be routinely detected by the standard FISH technique (Richards, Vogt, Muleris, Malfoy and Dutrillaux 1994), but in plants it is difficult to locate 10 kb

sequences (Guzzo, Campagnari and Levi 2000) even when amplification of FISH signals is applied to enhance high resolution FISH efficiency. As an alternative, large genomic clones such as lambda phages, cosmids, and bacterial artificial chromosomes (BAC), can be successfully used (Mukai 2005): the large amount of repeated sequences contained in them will facilitate homologous hybridization and signal detection.

Genome physical mapping is a centerpiece of structural, functional and comparative genomics research. Access to a genome with large fragments of DNA is currently essential for genome analysis. Cloning of large DNA fragments is possible with bacterial artificial chromosomes (BAC). Artificial chromosomes are laboratory constructs that contain DNA sequences and are used to introduce and control new DNA in a cell, to study how chromosomes function, and to map genes in genomes. This technique has been developed by Shizuya et al. (1992). BACs have been very useful in the development of genome libraries. Specific clones isolated from BAC libraries can be used for FISH mapping to generate high-density cytomolecular maps (Lichter et al. 1990). The use of genomic DNA cloned in large-insert vector BACs as probes in FISH experiments is called BAC-FISH. Hybridization of cloned BACs on metaphasic chromosomes allows their assignation to a particular chromosome, which is useful for physical mapping and chromosome identification. In cotton this technique has been used to assign unassigned linkage groups to specific chromosomes. Indeed, despite significant progress to construct genetic maps in the tetraploid cotton *G. hirsutum* (Reinisch, Dong, Brubaker, Stelly, Wendel and Paterson 1994; Rong, Abbey, Bowers, Brubaker, Chang, Chee, Delmonte, Ding, Garza, Marler, Park, Pierce, Rainey, Rastogi, Schulze, Trolinder, Wendel, Wilkins, Williams-Coplin, Wing, Wright, Zhao, Zhu and Paterson 2004; Nguyen, Giband, Brottier, Risterucci and Lacape 2004; Han, Wang, Song, Guo, Gou, Li, Chen and Zhang 2006), there remained six linkage groups that were not associated with specific chromosomes, which was a hindrance for integrated genetic map construction. These linkage groups were assigned to the A subgenome (A01, A02, and A03) or the D subgenome (D02, D03, and D08) by analysis of marker loci in two progenitor diploid species (*G. herbaceum* and *G. raimondii*) of the allotetraploid cotton (Reinisch, Dong, Brubaker, Stelly, Wendel and Paterson 1994). BAC-FISH have been used by Wang et al. (2006) to resolve this problem. Specific BAC clones constructed in *G. hirsutum* acc. TM-1 for these six linkage groups were identified by screening the BAC library using linkage group-specific simple-sequence repeats markers. These BAC clones were hybridized to ten translocation heterozygotes of *G. hirsutum* as BAC-fluorescence *in situ* hybridization probes. The results obtained allowed Wang et al. (2006) to assign the six unassigned linkage groups A01, A02, A03, D02, D03, and D08 to chromosomes 13, 8, 11, 21, 24, and 19, respectively, establishing the 13 homeologous chromosome pairs.

FISH techniques offer new potential not only for more reliable chromosome identification, but also regarding the integration of genetic and physical maps, for ordering molecular markers and measuring physical genome distances, and for structural and functional chromosome analyses. Molecular cytogenetic

markers are very important for plant genome analysis and genetic manipulation. To develop molecular cytogenetic markers in cotton, Zhang, Dong, Decanini, Lee, Ren, Yan, Kohel, Yu, Zhang and Stelly (2002) selected 8 BAC clones for FISH to somatic cells. Six of these 8 selected BAC clones yielded unambiguous signals on the distal region of A or D subgenomes.

3.5 Total Genomic DNA as Probe in FISH : GISH

Genomic *in situ* hybridization (GISH), a modification of FISH, allows chromosomes from different parents in hybrid plants to be painted in different colours. The technique uses total genomic DNA from one species as the labelled probe in hybridization experiments to chromosomal DNA. The addition of an excess amount of unlabelled DNA (blocking DNA) from the parent not used as a probe substantially increases the specificity of probing and enables more closely related species to be distinguished. The blocking DNA hybridizes to sequences in common between the block and the species used as labelled probe, so that mainly genome specific sequences remain exposed as sites for probe hybridization. In cells viewed under a fluorescence microscope, target DNA hybridized to the probe fluoresces differentially to the non-hybridized unrelated DNA sequences. Simultaneous multiple target-sequence detection (multicolor Genomic *in situ* hybridization) is now possible. GISH is a powerful tool that permits characterization of genomes and chromosomes in allopolyploid species. It can thus be used to differentiate the chromosomes of the different

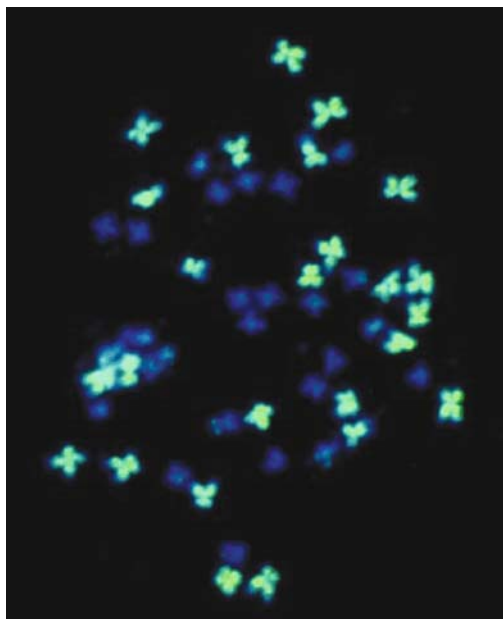


Fig. 3 Use of GISH to differentiate the A subgenome (in green) and the D subgenome (in blue) of *G. hirsutum*. Total DNA of *G. herbaceum* (A1 genome) labeled with digoxenin was used as probe; and total DNA of *G. thurberi* was used as blocking DNA (See Color Insert)

Fig. 4 Use of GISH to detect the presence of *G. longicalyx* chromosomes in mitotic plates of the trispecific hybrid [(*G. hirsutum* x *G. thurberi*)² x *G. longicalyx*]. Total DNA of *G. longicalyx* (labeled with digoxenin) and *G. hirsutum* (labeled with biotine) were used as probes. *G. longicalyx* genetic material is revealed in green (See Color Insert)

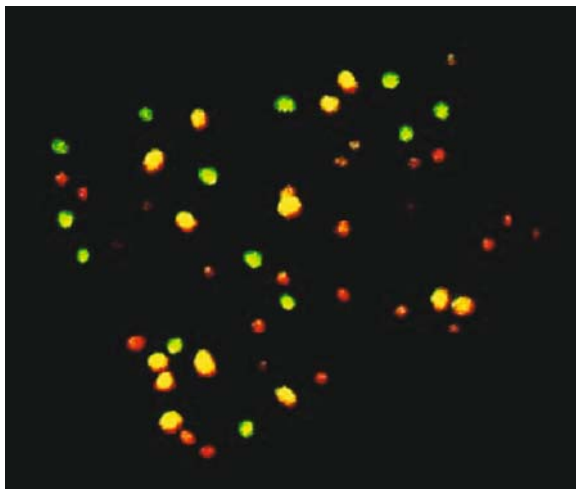
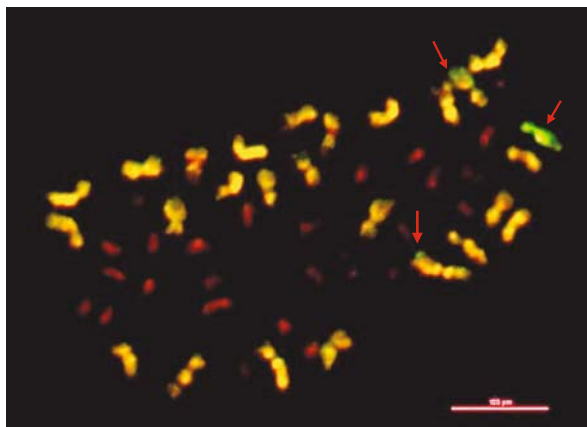


Fig. 5 Detection of recombinations (arrows) in a BC2 plant of the trispecies hybrid [(*G. hirsutum* x *G. thurberi*)² x *G. longicalyx*]. Total DNA of *G. longicalyx* (labeled with digoxenin) and *G. hirsutum* (labeled with biotine) were used as probes. *G. longicalyx* genetic material is revealed in green (See Color Insert)



subgenomes in tetraploid cottons (Fig. 3) and to recognize the genomic source of each chromosome involved in mitotic configurations of *Gossypium* interspecific hybrids (Fig. 4).

GISH can also be used to detect various structural alterations that follow hybrid and allopolyploid formation (Fig. 5).

4 Conclusion and Prospects

Classical cytogenetics played a fundamental role in revealing the taxonomy, chromosomal evolution, and phylogenetic relationships of cotton *species*. It also provided very efficient tools for rational exploitation of intraspecific,

interspecific, or induced variability in *Gossypium*. The main lessons that could be drawn from these contributions are presented in the first part of the present chapter.

Molecular cytogenetic techniques increase notably the speed, sensitivity and specificity of conventional cytogenetic methods. They also open new possibilities to study the structure of *Gossypium* genomes and chromosomal aberrations. Molecular cytogenetics allows the detection and more detailed localization of chromosomal rearrangements, both in mitotic and meiotic cytological preparations. With the aid of this new discipline, it is now possible to identify and molecularly characterize multiple repetitive sequences and localize them physically. Molecular cytogenetics also provides a novel chromosomal approach to complement DNA marker utilization, facilitating genome analysis to visualize genomic organization, chromosome structure, and landmarks for looking at genes, their clustering and orientation. The success of this novel approach in *Gossypium* is greatly enhanced by the availability of fundamental and practical knowledge gained through classical cytogenetic studies.

Molecular cytogenetic now allows work on extended DNA fibers (EDFs) from interphase nuclei and even from metaphase chromosomes (Lavania, Yamamoto and Mukai 2003). FISH on extended DNA fibers is a useful tool for determining the sizes of target DNA sequences, the order of genes or clones and their distances in a large chromosomal region (Fransz, Alonso-Blanco, Liharska, Peeters, Zabel and de Jong 1996; Suzuki, Tanaka, Yamamoto, Tomita, Kowyama and Mukai 2004). Such techniques can apply to *Gossypium* and should open immense possibilities in cotton genomic research. With the current state of technological advancement and future possibilities, cytogenetics appears as a powerful tool and an absolute must for cotton genetic study and exploitation in future years.

References

- Abraham, P. (1940) Cytological studies in *Gossypium*. I. Chromosome behaviour in the interspecific hybrid *G. arboreum* x *G. stocksii*. Indian J. Agric. Sci. 10, 285–298.
- Ahoton, L., Lacape, J.M., Baudoin, J.-P. and Mergeai, G. (2003) Introduction of Australian diploid cotton genetic variation into Upland Cotton. Crop Sci. 43, 1999–2005.
- Banerji, I. (1929) The chromosome numbers of Indian cotton. Ann. Bot. 43, 603–607.
- Baranov, P. (1930) Work of the cytoanatomical laboratory of N.I.Kh.I during the growing period of 1930. Bull. Sci. Inst. Cotton Cult., Tashkent 5, 7–17.
- Beal, J.M. (1928) A study of the heterotypic prophase in the microsporogenesis of cotton. Cellule 38, 247–268.
- Beasley, J.O. (1940) The production of polyploids in *Gossypium*. J. Hered. 31, 39–48.
- Beasley, J.O. (1942) Meiotic chromosome behavior in species, species hybrids, haploids and induced polyploids of *Gossypium*. Genetics 27, 25–53.
- Bergey, D.R., Stelly, D.M., Price, H.J. and McKnight, T.D. (1989) In situ hybridization of biotinylated DNA probes to cotton meiotic chromosomes. Stain Technol. 64, 25–37.
- Boza, B.T. and Madoo, R.M. (1941) Investigations on the relationship of the peruvian cotton species *G. raimondii* Ulb. Min. Fom. Direction de Agric. Ganaderia, Bol. 22, Lima, Peru, 29 p.

- Brown, M. S. (1950) Cotton from Bikini. *J. Hered.* 41, 115–121.
- Brown, M.S. (1980) The identification of the chromosomes of *Gossypium hirsutum* L. by means of translocations. *J. Hered.* 71, 266–274.
- Brown, M.S., and Menzel, M.Y. (1950) New trispecies hybrids in cotton. *J. Hered.* 41, 291–295.
- Brown, M.S. and Menzel, M.Y. (1952) The cytology and crossing behavior of *Gossypium gossypoides*. *Bulletin of the Torrey Botanical Club* 79, 110–125.
- Brown, M.S., Menzel, M.Y., Hasenkampf, C.A. and Naqi, S. (1981) Chromosome configurations and orientations in 58 heterozygous translocations in *Gossypium hirsutum*. *J. Hered.* 72, 161–168.
- Crane, C.F., Price, H.J., Stelly, D.M. and Czeschin, D.G. (1993) Identification of a homeologous chromosome pair by in situ DNA hybridization to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). *Genome* 36, 1015–1022.
- Davie, J.H. (1933) Cytological studies in the Malvaceae and certain related families. *J. Genet.* 28, 33–67.
- Demol, J. (1966) Amélioration de la résistance de la fibre chez *Gossypium hirsutum* L. Publication de l'Institut national pour l'étude agronomique au Congo Belge (Ineac), série scientifique, n°108. Bruxelles: Ineac.
- Denham, H.J. (1924) The cytology of the cotton plant. II. Chromosome numbers of old and new world cottons. *Ann. Bot.* 38, 433–438.
- Edwards, G.A. and Mirza, M.A. (1979) Genomes of the Australian wild species of cotton. II. The designation of a new G genome for *Gossypium bickii*. *Can. J. Genet. Cytol.* 21, 367–372.
- Endrizzi, J.E. (1957) The cytology of two hybrids of *Gossypium*. *J. Hered.* 48, 221–226.
- Endrizzi, J.E., Turcotte, E.L. and Kohel, R.J. (1985) Genetics, cytogenetics, and evolution of *Gossypium*. *Adv. Genet.* 23, 271–375.
- Franz, P.F., Alonso-Blanco, C., Liharska, T.B., Peeters, A.J.M., Zabel, P. and de Jong, J.H. (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibers. *Plant J.* 9, 421–430.
- Gerstel, D.U. (1953) Chromosome translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* 7, 234–244.
- Guzzo, F., Campagnari, E. and Levi, M. (2000) A new FISH protocol with increased sensitivity for physical mapping with short probes in plants. *Journal of Experimental Botany* 51, 965–970.
- Han, Z.G., Wang, C.B., Song, X.L., Guo, W.Z., Gou, J.Y., Li, C.H., Chen, X.Y. and Zhang, T.Z. (2006) Characteristics, development and mapping of *G.hirsutum* derived-EST-SSRs in allotetraploid cotton. *Theor Appl Genet* 112, 430–439.
- Hanson, R.E., Islam-Faridi, M.N., Percival, E.A., Crane, C.F., Ji, Y., McKnight, T.D., Stelly, D.M. and Price, H.J. (1996) Distribution of 5S and 18S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. *Chromosoma* 105, 55–61.
- Harland, S.C. (1928) Cotton notes. *Trop. Agric.* 5, 116–117.
- Hau, B. (1981) Lignées d'addition sur l'espèce *G. hirsutum* L. I. Utilisation de l'hybridation interspécifique et de la méthode des lignées d'addition pour l'amélioration du cotonnier. *Coton Fibres Trop.* 36, 247–258.
- Iyengar, N.K. (1944). Cytological investigations on hexaploid cottons. *Indian J. Agric. Sci.* 14, 142–151.
- Ji, Y.F., De Donato, M., Cranel, C.F., Raska, W.A., Islam-Faridi, M.N., McKnight, T.D., Price, H.J. and Stelly, D.M. (1999) New ribosomal RNA gene locations in *Gossypium hirsutum* mapped by meiotic FISH. *Chromosoma* 108, 200–207.
- Ji, Y., Raska, D.A., McKnight, T.D., Islam-Faridi, N.M., Crane, C.F., Zwick, M.S., Hanson, R.E., Price, H.J. and Stelly, D.M. (1997) Use of meiotic fluorescence in situ hybridization for identification of a new monosome in *Gossypium hirsutum* L. *Genome* 40, 34–40.
- Ji, Y., Raska, W. A., De Donato, M., Islam-Faridi, M. N., Price, H.J. and Stelly, D. M. (1999) Identification and distinction among segmental duplication-deficiencies by fluorescence in situ hybridization (FISH)-adorned multivalent analysis. *Genome* 42, 763–771.

- Jiang, J. and Gill, B.S. (1994) New 18S–26S ribosomal RNA gene loci: chromosomal landmarks for the evolution of polyploid wheats. *Chromosoma* 103, 179–185.
- Kammacher, P. (1960) Observations cytologiques sur deux hybrides F1 entre espèces cultivées de cotonnier et l'espèce diploïde sauvage *Gossypium raimondii* Ulb. *Rev. Cytol. Biol. Vég.* 22, 1–32.
- Katterman, F.R.H. and Ergle, D.R. (1970) A study of quantitative variations of nucleic acids in *Gossypium*. *Phytochemistry* 9, 2007–2010.
- Kearney, T.H. (1930) Cotton plants, name and wild and genetics of cotton. *J. Hered.* 21, 325–336.
- Kimber, G. (1961) Basis of the diploid-like meiotic behavior of polyploid cotton. *Nature* 191, 98–100.
- Konan, O.N., D'Hont, A., Baudoin, J.-P. and Mergeai, G. (2007) Cytogenetics of a new trispecies hybrid in cotton: [(*Gossypium hirsutum* L. x *G. thurberi* Tod.)² x *G. longicalyx* Hutch. & Lee]. *Plant Breeding* 126, 176–181.
- Lavana, U.C., Yamamoto, M. and Mukai, Y. (2003) Extended chromatin and DNA fibers from active plant nuclei for high resolution FISH. *J. Histochem. Cytochem.* 51, 1249–1253.
- Longley, A.E. (1933) Chromosomes of *Gossypium* and related genera. *J. Agric. Res.* 46, 217–227.
- Louant, B.P. and Maréchal, R. (1975) Comportement méiotique des hybrides trispécifiques (*Gossypium thurberi* Tod. x *G. anomalum* Wawr.) doublé x *G. hirsutum* L. et (*G. hirsutum* L. x *G. anomalum* Wawr.) doublé x *G. harknessii* Brandg. *Coton et Fibres Trop.* 30, 383–387.
- Louant, B.P., Maréchal, R. and Baudoin, J.-P. (1977). Les facteurs agissant sur les transferts de matériel génétique chez les hybrides bispécifiques : Principes utiles aux sélectionneurs et perspectives dérivant de l'étude des croisements interspécifiques dans le genre *Gossypium*. *Coton Fibres Trop.* 32, 39–57.
- Maréchal, R. (1972) Comportement méiotique chez l'hybride *Gossypium hirsutum* x *G. areysianum* (Delf.) Hutch. aux niveaux triploïde, hexaploïde et pentaploïde. *Coton et Fibre Trop.* 27, 205–211.
- Maréchal, R. (1974) Analyses de la conjugaison méiotique chez les hybrides triploïdes entre *Gossypium hirsutum* L. et des espèces sauvages australiennes. *Bull. Rech. Agron. Gem-boux* 9, 193–204.
- Menzel, M.Y. (1955) A cytological method for genome analysis in *Gossypium*. *Genetics* 40, 214–223.
- Menzel, M.Y. and Brown, M.S. (1954) The significance of multivalent formation in three-species *Gossypium* hybrids. *Genetics* 39, 546–557.
- Menzel, M.Y. and Brown, M.S. (1978a) Genetic lengths and break points in twelve chromosomes of *Gossypium hirsutum* involved in ten reciprocal translocations. *Genetics* 88, 541–558.
- Menzel, M.Y. and Brown, M.S. (1978b) Reciprocal chromosome translocations in *Gossypium hirsutum*. Arm locations of breakpoints and recovery of duplication-deficiencies. *J. Hered.* 69, 383–390.
- Menzel, M.Y. and Richmond, K.L. (1985) A revised map of chromosome 15 in *Gossypium hirsutum*. *Can. J. Genet. Cytol.* 28, 272–277.
- Menzel, M.Y. and Dougherty, B.J. (1987) Transmission of duplication-deficiencies from cotton translocations is unrelated to map lengths of the unbalanced segments. *Genetics* 116, 321–330.
- Menzel, M.Y., Richmond, K.L. and Dougherty, B.J. (1987) A revised map of chromosome 16 in *Gossypium hirsutum*. *Genome* 29, 823–827.
- Menzel, M.Y., Richmond, K.L., and Dougherty, B.J. (1985) A chromosome translocation breakpoint map of the *Gossypium hirsutum* genome. *J. Hered.* 76, 406–414.
- Meredith, W.R. and Bridge, R.R. (1971) Break up of linkage blocks in cotton *Gossypium hirsutum* L. *Crop Sci.* 11, 695–698.

- Mergeai, G., Baudoin, J.P. and Vroh Bi, I. (1997) Exploitation of trispecific hybrids to introgress the glandless seed and glanded plant trait of *Gossypium sturtianum* Willis in *G. hirsutum* L. Biotech. Agron. Soc. Environ. 1, 272–277.
- Miller, P.A. and Rwalings, J.O. (1967) Break up of initial linkage blocks through intermating in a cotton breeding population. Crop Sci. 7, 199–204.
- Mukai, Y. (2005) Perspectives in molecular cytogenetics of wheat. Wheat Inf. Serv. 100, 17–31.
- Mukai, Y., Friebe, B. and Gill, B.S. (1992) Comparison of C-banding patterns and *in situ* hybridization sites using highly repetitive and total genomic rye DNA probes of 'Imperial' rye chromosomes added to 'Chinese Spring' wheat. Japan J. Genet. 67, 71–83.
- Murovenko, O.V., Fedotov, A.R., Punina, E. O., Fedorova, L.I., Valerii, G.G. and Zelenin, A.V. (1998) Comprison of chromosome BrdU-Hoechst-Giemsa banding patterns of the A₁ and (AD)₂ genomes of cotton. Genome 41, 616–625.
- NDungo, V., Demol, J. and Maréchal, R. (1988a) L'amélioration du cotonnier *Gossypium hirsutum* L. par hybridation interspécifique. 1. Phylogénie et speciation du genre *Gossypium*. Bull. Rech. Agron. Gembloux 23, 27–49.
- NDungo, V., Demol, J. and Maréchal, R. (1988b) L'amélioration du cotonnier *Gossypium hirsutum* L. par hybridation interspécifique. 2. Méthodologie pour l'exploitation de la diversité génétique du genre *Gossypium*. Bull. Rech. Agron. Gembloux 23, 171–204.
- NDungo, V., Demol, J. and Maréchal, R. (1988c) L'amélioration du cotonnier *Gossypium hirsutum* L. par hybridation interspécifique. 3. Application et résultats obtenus. Bull. Rech. Agron. Gembloux 23, 283–316.
- Nguyen, T.B., Giband, M., Brottier, P., Risterucci, A.M. and Lacape, J.M. (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109, 167–175.
- Nikolajeva, A. (1923) A hybrid between Asiatic and American cotton plant *Gossypium herbaceum* L. and *Gossypium hirsutum* L. Bull. Appl. Bot. Plant Breed. 13, 117–134.
- Phillips, L.L., and Strickland, M.A. (1966) The cytology of a hybrid between *Gossypium hirsutum* and *G. longicalyx*. Can. J. Genet. Cytol. 8, 91–95.
- Poisson, C. (1970) Contribution à l'étude de l'hybridation interspécifique dans le genre *Gossypium*: transfert de matériel génétique de l'espèce sauvage diploïde *Gossypium anomalum* à l'espèce cultivée tétraploïde *G. hirsutum*. Doctorat es Sciences, Univ. Paris Sud, Centre d'Orsay, 76 p.
- Price, H.J., Stelly, D.M., McKnight, T.D., Scheuring, C.F., Raska, D., Michaelson, M.J. and Bergey, D. (1990) Molecular cytogenetic mapping of a nucleolar organizer region in cotton. J. Hered. 81, 365–370.
- Rayburn, A.L. and Gill, B.S. (1986) Molecular analysis of the D-genome chromosomes of wheat. J. Hered. 77, 253–255.
- Reinisch, A.J., Dong, J.M., Brubaker, C.L., Stelly, D.M., Wendel, J.F. and Paterson, A.H. (1994) A detailed RFLP map of cotton (*Gossypium hirsutum* x *Gossypium barbadense*): chromosome organization and evolution in a disomic polyploid genome. Genetics 138, 829–847.
- Richards, F., Vogt, N., Muleris, M., Malfoy, B. and Dutrillaux B. (1994) Increased FISH efficiency using APC probes generated by direct incorporation of labelled nucleotides by PCR. Cytogenetics and Cell Genetics 65, 169–171.
- Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behavior of hexaploid wheat. Nature 182, 713–715.
- Riley, R., Chapman, V. and Kimber, G. (1960) Position of the gene determining the diploid-like meiotic behavior of wheat. Nature 186, 259–260.
- Rong, J.K., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C., Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schulze, S.R., Trolinder, N.L., Wendel, J.F., Wilkins, T.A., Williams-Coplin, T.D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. and Paterson, A.H. (2004) A 3347-locus genetic

- recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166, 389–417
- Schwendiman, J., Koto, E. and Hau, B. (1980) Considérations sur l'évolution de l'appariement chromosomique chez les allotétraploïdes de cotonnier *G. hirsutum* x *G. stocksii* et *G. hirsutum* x *G. longicalyx* et sur la position taxonomique de *G. longicalyx*. *Coton Fibres Trop.* 35, 269–275.
- Sears, E.R. and Okamoto, M. (1958) Intergenomic chromosome relationships in hexaploid wheat. *Proc. 10th Intern. Congr. Genet.*, Montreal, Montreal, 2, 258–259.
- Skovsted, A. (1933) Cytological studies in cotton. I. The mitosis and the meiosis in diploid and triploid Asiatic cotton. *Ann. Bot.* 47, 227–258.
- Skovsted, A. (1934a) Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet.* 28, 407–424.
- Skovsted, A. (1934b) Cytogenetics in relation to plant breeding in cotton. *Rep. Summ. Proc. 2nd Conf. Cotton Grow. Probl., Emp. Cotton Grow. Corp.*, 46–51.
- Skovsted, A. (1935a) Cytological studies in cotton. III. A hybrid between *Gossypium davidsonii* Kell. and *G. sturtii* F. Muell. *J. Genet.* 30, 397–405.
- Skovsted, A. (1935b) Some new interspecific hybrids in the genus *Gossypium* L. *J. Genet.* 30, 447–463.
- Skovsted, A. (1937) Cytological studies in cotton. IV. Chromosome conjugation in interspecific hybrids. *J. Genet.* 34, 97–134.
- Stephens, S.G. (1947) Cytogenetics of *Gossypium* and the problem of the origin of New World cottons. *Adv. Genet.* 1, 431–442.
- Stewart, J.M. (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In : G.A. Constable and N.W. Forrester (eds), *Proc. World Cotton Research Conference-1. Challenging the future*. Brisbane, Australia, February 14–17, 1994, CSIRO, Melbourne, 313–327.
- Shuijin, Z. and Biling, L. (1993) Studies of introgression of the glandless seeds-glanded plant trait from *Gossypium bickii* into cultivated upland cotton (*G. hirsutum*). *Cotton Fibres Trop* 43, 195–199.
- Suzuki, G., Tanaka, S., Yamamoto, M., Tomita, R.N., Kowayama, Y. and Mukai, Y. (2004) Visualization of the S-locus region in *Ipomoea trifida*: toward positional cloning of self-incompatibility genes. *Chromosome Res.* 12, 475–481.
- Tsujimoto, H., Mukai, Y., Akagawa, K., Nagaki, K., Fujigaki, J., Yamamoto, M. and Sasakuma, T. (1997) Identification of barley chromosomes by repetitive sequences: Conservative distribution of Afa-family repetitive sequences on the chromosomes of barley and wheat. *Genes Genet. Syst.* 72, 303–309.
- Vroh Bi, I., Baudoin, J.-P., Hau, B. and Mergeai, G. (1999) Development of high-gossypol cotton plants with low gossypol seeds using trispecies bridge crosses and in vitro culture of seed embryos. *Euphytica* 106, 243–51.
- Wang, K., Song, X., Han, Z., Guo, W., Yu, J.Z., Sun, J., Pan, J., Kohel, R.J. and Zhang, T. (2006) Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping. *Theor. Appl. Genet.* 113, 73–80.
- Webber, J.M. (1934a) Cytogenetic notes on cotton and cotton relatives. *Science* 80, 268–269.
- Webber, J.M. (1934b) Chromosome number and meiotic behavior in *Gossypium*. *J. Agric. Res.* 49, 223–237.
- Webber, J.M. (1935) Interspecific hybridization in *Gossypium* and the behavior of F1 plants. *J. Agric. Res.* 51, 1047–1070.
- Webber, J.M. (1939) Relationships in the genus *Gossypium* as indicated by cytological data. *J. Agric. Res.* 58, 237–261.
- Wendel, J.F. and Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78, 139–186.
- Woo, S.-S. (1996) Development of a map-based cloning system in *Sorghum bicolor*:
1) Isolation of megabase-size DNA and construction of a bacterial artificial

- chromosome library and 2) genetic and physical mapping of the 5S rDNA locus. Ph.D. dissertation, Texas A&M University, College Station, Tex.
- Wouters, W. (1948) Contribution à l'étude taxonomique et caryologique du genre *Gossypium* et application à l'amélioration du cotonnier au Congo Belge. Publications de l'I.N.E.A.C. Serie Scientifique 34, 1–383.
- Youngman, W. and Pande, S.C. (1927) Occurrence of branched hairs in cotton and upon *Gossypium stockssii*. Nature (London) 119, 845.
- Zhang, L.T., Dong, J., Decanini, L.I., Lee, M.-K., Ren, C., Yan, B., Kohel, R.J., Yu, J., Zhang, H. and Stelly, D.M. (2002) Development of molecular cytogenetic markers in cotton. Plant, Animal & Microbe Genomes X Conference, January 12–16, 2002 Town & Country Convention Center San Diego, CA.

Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development

Peng W. Chee and B. Todd Campbell

Abstract Cotton is the single most important natural fiber in the world and represents a vital agricultural commodity in the global economy. Ninety percent of cotton's value resides in the lint fiber. Cotton fiber quality, defined by the physical properties of the lint fibers, is an important part of the cotton manufacturing process from field harvest through ginning and textile manufacturing and is reflected in the end product. The primary fiber properties affecting textile manufacturing and end product quality include fiber length and uniformity, strength, elongation, fineness, and maturity. Numerous techniques and tools to measure these fiber properties have been developed during the last 100 years. Classical quantitative genetics research methods have determined the heritability, components of genetic variance, environmental interactions, and correlations of fiber properties among one another and with fiber yield. In response to the advances made in fiber processing and manufacturing over the course of the 20th and 21st centuries, classical plant breeding based on phenotypic selection has improved fiber quality while also increasing fiber yields. At the same time, intensive phenotypic selection programs have resulted in decreased levels of genetic diversity within the primary gene pool of Upland cotton. Classical plant breeding programs have faced challenges and difficulties transferring new, stably inherited allelic variation from inter-specific hybridization. However, the last 15 years have witnessed an explosion of efforts to utilize molecular biology tools to study the structure, function, and evolutionary relationships of the cotton genome. Much of the first 15 years of molecular genetic research into cotton fiber quality has been devoted to developing core infrastructure including polymorphic DNA markers, discrete genetic mapping populations, and extensive nuclear genetic linkage maps. This activity has provided insight into the location, effects, and complexity of the quantitative trait loci (QTL) associated with fiber properties. A fascinating story is being written from the advances being made by combining classical and molecular

P.W. Chee (✉)

Department of Crop and Soil Sciences, University of Georgia
e-mail: pwchee@uga.edu

genetics to explore fiber quality. Although fiber properties are affected by a large number of small effect QTLs, molecular research has also demonstrated that a large percentage of the loci controlling fiber quality properties are present in “gene islands” that are non-randomly distributed across the A- and D-genomes. The next 15 years of molecular genetic research will undoubtedly provide a clearer picture of the genetic basis of cotton fiber quality and the functions of genes controlling various fiber properties. Future research efforts that combine the power of molecular genetics with the knowledge and experience accrued by classical plant breeding will provide portable and inexpensive DNA markers that can be used by plant breeders to select and develop the next generation of high fiber quality cotton cultivars.

1 Introduction

Cotton is unique in that four different species in the genus *Gossypium* (Malvaceae) were independently domesticated in two separate continents for lint fiber production (Wendel and Cronn 2003). Therefore, the word “cotton” in the textile industry can apply equally to the two tetraploid species *G. hirsutum* (Upland) and *G. barbadense* (Pima, Sea Island, or Egyptian) endemic to the Americas, as well as the two diploid species *G. arboreum* and *G. herbaceum* endemic to Africa and Asia. Currently, the two tetraploid species supply an overwhelming majority of the world textile fiber commerce, with *G. hirsutum* alone responsible for over 90% of the world’s total fiber production.

The epidermal layer of cottonseed contains two distinct types of fibers. The short fibers, called linters or fuzz, are tightly adhered to the seed coat. The long fibers, called lint, are loosely adhered to the seed and readily removed during the ginning process. The fuzz fiber represents an important source of raw material for the manufacture of paper and other industrial cellulose products. However the lint fibers, which are longer and spinnable, are the natural cellulose products that make cotton the world’s most important fiber crop. Worldwide cotton fiber production has now reached 115 million bales, and contributes about 40% of the world fiber market (ICAC 2005), making cotton the single most important natural fiber in textile industries and a vital agricultural commodity in the global economy. In 2005, the aggregate value of the world’s cotton crop was estimated to be about US\$30 billion/yr, with 90% of its value resides in the lint fiber.

Traditionally, genetic improvement of cotton fiber quality has relied on quantitative genetics and phenotypic selection through field-based plant breeding. In response to the advances made in fiber processing and manufacturing over the course of the 20th century to date, phenotypic selection based plant breeding has certainly improved fiber quality while also increasing fiber yields. However, the last 15 years have witnessed an explosion of efforts to utilize molecular biology tools to study the structure, function and evolutionary relationships of the cotton genome. Central to the genomic toolkit is the

molecular map of the cotton genome, which has provided a common point of reference for dissecting the quantitative variation for a host of phenotypes, including those related to cotton fiber. Collectively, these molecular genetic investigations have greatly extended our understanding of the quantitative trait loci (QTL) influencing fiber quality such as fiber length, strength, fineness and elongation. In view of their dominance in fiber production, the bulk of the research on fiber quality QTL discovery has been conducted on the two cultivated tetraploid species. In this chapter, we seek to summarize the current status of molecular genetics applied to cotton fiber quality. In addition, we will discuss the implications of the new tools and knowledge gained through our recent understanding of the cotton genome with their potential application to further improve cotton fiber quality.

2 What is Fiber Quality and How is it Measured?

Cotton fiber quality is defined by the physical properties of the individual fibers. Lint fiber is generally spun into yarn, which is then weaved or knitted into different types of fabrics as dictated by its quality and the desired properties of the end product. Fiber quality, therefore, is the combination of fiber properties that affect the efficiency of yarn-spinning, weaving and other fabric manufacturing processes as well as the quality associated with cotton fabrics. Each step in the manufacturing process ultimately contributes to the function and utilization of textile products. Currently, there are three major types of yarn-spinning technologies that include rotor, air-jet, and ring systems (May 2000). Unfortunately, each type of spinning technology requires a different combination of fiber quality properties to operate most efficiently. Ring spinning is facilitated by minimum fiber length, high fiber strength, and fine fiber. Rotor spinning is facilitated by high fiber strength and fineness ahead of fiber length. Air-jet spinning is facilitated by uniform fiber length and low short fiber content. The importance of these inherent fiber properties is such that the textile industry has invested extensively in devising instruments to measure the physical properties of fibers and to evaluate their significance for processing and textile uses.

Each individual bale of cotton produced and marketed in the United States and the World is subjected to fiber property classification. Although numerous instruments have been developed and used over the 20th century to measure fiber properties, the classification of US cotton is carried out using the High Volume Instrument (HVI) by the US Department of Agriculture. The standard HVI currently consists of determinations of fiber length, length uniformity, strength, micronaire, and color; while subjective determinations of leaf grade, preparation, and extraneous matter are recorded by trained classing personnel. In recent years, the Advanced Fiber Information System (AFIS) instrument (Uster Technologies AG, Memphis, TN) has been developed to measure a series of fiber properties as an alternative to HVI based fiber property measures. Research studies relating to

fiber quality properties have utilized the HVI, AFIS, and additional instruments to measure specific fiber properties. Overall, the main fiber properties highly correlated with spinning performance and end product quality include those related to the length, strength, elongation and fineness of the fiber. These fiber properties are being used to determine the price premiums and discounts associated with the market value of raw cotton. The importance of the main fiber properties and how they are measured are discussed below.

2.1 Fiber Length and Length Distribution

The fiber length and length distribution of raw cotton ultimately determines the end-product produced through the spinning and manufacturing process. Fiber length, as determined by HVI, is the average length of the longer one-half of the fibers by weight and is also known as the upper half mean length. Other instruments used to measure length include the fibrograph (Hertel 1940) and the AFIS. The fibrograph provides length measurements that measure the distance required to span 2.5% and 50% of a 'beard' or preparation of cotton fibers. The AFIS instrument provides a series of length measurements on individual fibers including length by weight, length by number, and upper quartile length by weight. Knowledge of fiber length is necessary to manufacture a yarn of specific size on ring spinning systems, and typically longer fibers are used to manufacture fine yarns (May 2000). Length distribution is also an important consideration during the spinning process when fiber lengths are sufficiently long to meet yarn-spinning requirements (Behery 1993; May 2000). Hence, the distribution of fiber length becomes increasingly important for optimum spinning efficiency if fibers are sufficiently long. Currently, a direct measure of length distribution is not available; however, the HVI measures fiber length uniformity and predicts short fiber content. HVI fiber length uniformity is calculated from a ratio of the HVI mean length and upper half mean lengths of the fibers and is expressed as a percentage. A length uniformity ratio can also be calculated from the ratio of fibrograph 2.5% and 50% span length measurements. HVI short fiber content is predicted from the upper half mean length and length uniformity using multiple regression procedures. The AFIS instrument can also be used to measure length uniformity and short fiber content. AFIS length uniformity is assessed by calculating a coefficient of variation for length by number and length by weight measures. AFIS short fiber content is obtained by measuring the amount of short fibers by weight and/or number that are below 12.7 mm long.

2.2 Fiber Strength

The fiber strength of raw cotton is highly correlated with yarn strength and directly impacts the production speed of woven or knitted fabric. Fiber strength

ultimately impacts the fiber's durability throughout the harvest, ginning, and yarn manufacturing processes (May 2000). Fiber strength is also essential to maintain cotton's natural qualities after chemical processing of fabric. Overall, fiber strength is generally determined by the amount of force required to break a prepared bundle of fibers. There are several instruments capable of measuring fiber strength using this approach that includes the Pressley tester (Pressley 1942), Stelometer (Hertel 1953), and the HVI.

2.3 Fiber Elongation

Fiber elongation measures the elasticity of fibers before a break occurs (May 2000). Fiber elongation can be measured using a Stelometer or HVI and corresponds to the distance a bundle of fibers is stretched prior to break, expressed as a percentage. Fiber breakage causes inefficiency in yarn manufacturing performance and decreased end product quality. However, there are differences in opinion regarding how elongation correlates with yarn spinning performance. Backe (1996) reported that increased fiber elongation was positively correlated to spinning performance. On the contrary, Meredith (1991), Green and Culp (1990), and May and Taylor (1998) reported a negative correlation between fiber elongation and spinning performance. More recently, Benzina, Hequet, Abidi, Gannaway, Drean, and Harzallah (2007) reported a negative correlation with spinning performance and suggested that decreased fiber elongation resulted in optimum yarn strength and spinning performance. Until the relationship between fiber elongation and spinning performance is resolved and confirmed, plant breeding programs will likely continue to base selections on alternative fiber properties.

2.4 Fineness and Maturity

Finer and fully-mature raw cotton fibers can be spun into yarn with stronger fiber characteristics that affect spinning efficiency and end-product quality (Deussen 1992). Fineness and maturity ultimately affect the ability of manufacturers to produce products with the finest yarn counts. The most commonly used measure of fiber fineness and maturity is called micronaire, which is a module option of the HVI. Micronaire is measured as the resistance to airflow of a constant weight of fibers at a single air pressure (Johnson 1952). Unfortunately, the micronaire value is confounded by both fineness and maturity, unless one has knowledge of fineness or maturity (May 2000). Low micronaire cotton could result from immature and/or very fine fibers, properties which cannot be distinguished from one another. Other instruments that more precisely measure fineness or maturity include the Arealometer (Hertel and Craven 1951) and the Shirley fineness-maturity tester (American Society for Testing and Materials, 1993). The AFIS also provides a module that directly measures the diameter of individual fibers, thus providing a measure of fineness.

3 Heritability, Genetic Variation, and Improvement of Fiber Quality Properties

To maximize the probability of genetic improvement for a specific trait of interest, there are three general requirements that must be met. First, there must be a satisfactory methodology(s) developed to measure the trait of interest. Second, the trait of interest and its measurement must be heritable. Third, there must exist sufficient genetic variability to ultimately realize long-term genetic improvement for the trait of interest through plant breeding (Baenziger, Russell, Graef, and Campbell 2007). In the context of cotton fiber quality properties and their genetic improvement, the extent of the three general requirements to maximize genetic improvement of cotton fiber quality can be examined. Substantial scientific and engineering research, conducted since the early days of dedicated plant breeding for cotton improvement, has developed readily available instrumentation and methodology to measure specific fiber properties on a small scale.

Historically, classical genetic studies over the last 50+ years, have tested the feasibility of using these fiber quality instruments to measure and improve specific fiber properties. Most studies have evaluated the genetic basis and heritability of specific fiber properties within the primary gene pool of tetraploid cotton by studying genetic populations derived from intra-specific and inter-specific crosses involving *G. hirsutum* and/or *G. barbadense*. The bulk of these genetic studies have documented the heritable nature of fiber properties, thus indicating the ability to improve specific fiber properties through traditional hybridization, phenotypic selection, and various plant breeding approaches.

Table 1 provides a summary of heritability studies involving fiber properties revised from May (2000). The mean broad-sense heritability, or proportion of phenotypic variance explained by genetic variance, provides evidence that fiber properties are amenable to improvement through selection. Moreover, the mean narrow-sense heritability, or proportion of phenotypic variance explained by additive genetic variance, provides evidence that fiber properties can be improved using inbreeding and pedigree selection methods. Although heritability estimates for fiber properties are moderate to high, the inheritance of fiber properties appears to be quantitative in nature and controlled by multiple genes. Trait histograms for different segregating populations have consistently shown normal distributions for fiber length, strength, elongation, and fineness. However, there are reports indicating fiber strength is controlled by 2–3 major genes (Meredith 2005). Genotype-by-environment ($G \times E$) interactions for fiber quality properties have also been examined in numerous studies. Information concerning the extent of $G \times E$ interactions is critically important for cotton breeding programs to recognize a given cultivar's likely area of adaptation (Campbell and Jones 2005). Meredith (1984) summarized the size of $G \times E$ interaction components relative to the genotype component in

Table 1 Heritability estimates for fiber quality traits

Trait	Broad-sense		Narrow-sense	
	Mean	Range	Mean	Range
Length	0.77	0.54—0.88	0.55	0.10—1.00
Strength	0.64	0.15—0.90	0.57	0.10—0.86
Elongation	0.59	0.21—0.80	0.60	0.36—0.90
Micronaire	0.71	0.53—0.87	0.33	0.08—0.53
Fineness	0.67	0.61—0.72	0.47	0.07—0.68

Data are summarized from May (2000) and were derived from a divergent sample of genetic populations, various selection units, and test instruments.

a range of studies and concluded that fiber property $G \times E$ interactions were generally small for fiber strength and length, but larger for fineness. More recently, Campbell and Jones (2005) examined HVI derived fiber quality properties and found that 8–24 % of the total variation could be attributed to $G \times E$ interactions. They also reported that $G \times E$ interactions were generally due to changes in magnitude rather than changes in the rank of genotypes.

Genetic variation for fiber quality properties and the components of genetic variance have also been studied extensively. Generally speaking, genetic variance can be partitioned into additive (additivity) and non-additive (dominance and epistasis) components. Quantitative genetic models can also be extended to include the interactions of each genetic component with the environment. The majority of studies of gene action have focused on additive and dominance components of genetic variance and generally assumed no epistasis. One way to evaluate gene action is to calculate and compare the additive and non-additive components of genetic variance. Table 2 provides a summary of the ratio of additive to non-additive genetic variance calculated in a series of experiments initially summarized by May (2000) and expanded here to include Tang, Jenkins, McCarty, and Watson (1997), Meredith and Brown (1998), McCarty, Jenkins, and Wu (2004), Cheatham, Jenkins, McCarty, Watson, and Wu (2003), Yuan, Zhang, Guo, Pan, and Kohel. (2005), and Mei, Syed, Gao, Thaxton, Smith, Stelly, and Chen (2007). Generally, genetic variation for fiber quality properties is predominantly controlled by additive gene action

Table 2 Ratio of additive to non-additive genetic variance for fiber quality traits

Trait	Mean	Range
Length	7.7	<0.1—48.0
Strength	46.8	<0.1—680.0
Elongation	18.0	<0.1—425.0
Fineness	7.4	<0.1—37.0

Data are summarized from May (2000) and additional literature to date that include a divergent sample of genetic populations, various selection units and test instruments.

rather than non-additive genetic variation. Additive genetic variance represents the majority of the total genetic variance for fiber strength and indicates the high propensity to improve fiber strength through phenotypic selection based on HVI or other strength measurements. Compared to fiber strength a larger percentage of the total genetic variance for fiber length and fineness is attributed to non-additive genetic variance. However, the greater preponderance of additive variance for both fiber properties also provides evidence that phenotypic selection can be used to improve length and fineness. The large ratio of additive to non-additive genetic variance found for fiber elongation also provides evidence that phenotypic selection can be used to improve fiber elongation. However, because a clear and repeatable relationship between fiber elongation and spinning performance is not well established, phenotypic selection for fiber elongation based on current elongation measures has not been a focus for improving fiber quality. Future fiber quality gene action studies should explore the effect of epistasis on genetic variance.

Although fiber properties are heritable and show additive genetic variance, the relationships between fiber properties and other traits of interest are also important to consider. If genetic correlations between two or more traits are high, the selection of one trait will simultaneously result in changes of the correlated trait(s) (Meredith 1984b). Genetic correlations between traits can be caused by linkage and/or pleiotropy. Early studies summarized by Meredith (1984a), reported that lint yield is negatively correlated with fiber strength and fiber length, while positively correlated with fiber elongation and micronaire. Overall, the relationship between lint yield and fiber strength has received the most attention. Subsequently, Culp, Harrell and Kerr (1979) reported a breaking of the negative relationship between fiber strength and lint yield by developing germplasm with high strength and lint yield, thus suggesting linkage as the causal agent for the negative relationship. Smith and Coyle (1997) also reported a similar negative relationship between lint percent and/or lint yield and fiber strength, but did not rule out pleiotropy as the causal agent. In addition to correlations with lint yield and lint percent, fiber properties also show correlations among one another. Table 3 provides a summary of genetic correlations among several fiber properties summarized from May (2000). Fiber length and strength are generally positively correlated. The range of correlations between fiber length, fineness, and elongation for the summarized studies do not provide definitive evidence that a strong correlation exists among these properties. Kloth (1998) also examined genetic correlations among fiber

Table 3 Summary of genetic correlations between fiber quality traits

Trait	Strength	Fineness	Elongation
Length	0.23 (–0.10–0.41)	0.11 (–0.48–0.70)	0.02 (–0.17–0.15)
Strength	–	–0.11 (–0.25–0.14)	–0.41 (–0.84–0.03)

Data are summarized from May (2000) that include a divergent sample of genetic populations, various selection units and test instruments. Range of correlation is given in parenthesis.

properties and confirmed the positive correlation between fiber strength and length. They also found positive correlations between fiber strength and maturity, as well as fineness and fiber elongation. Negative correlations were found between micronaire, fineness and fiber length and between fiber elongation, fineness and fiber strength. Overall, the consistent positive correlation between fiber strength and length among numerous studies suggest the ability to select for both properties simultaneously without deleterious effects on either trait.

Capitalizing on increased genetic knowledge of fiber properties and other traits of interest, plant breeders have practiced formal crossing and selection programs to develop cotton cultivars with superior lint yield and fiber quality since the beginning of the 20th century. Inbreeding, selection, and testing techniques have capitalized on accumulating the additive effects of new allele combinations (Campbell, Bowman, and Weaver 2008). Since fiber properties are predominantly under additive gene action, cultivar development and germ-plasm enhancement programs have successfully increased fiber quality by practicing phenotypic selection for fiber properties in early generations (May 2000). Numerous reports have documented the lint yield increases obtained over time through inbreeding and phenotypic selection with some reference to fiber quality. Positive genetic gains in lint yield have been reported for Upland, Acala, and Pima cotton breeding programs (Bridge, Meredith, and Chism 1971; Hoskinson and Stewart 1977; Bridge and Meredith 1983; Bassett and Hyer 1985; Culp and Green 1992; Zhang, Lu, Adragna, and Hughs 2005). These studies also reported an increase or at least maintenance of fiber quality over this same period of time. Genetic gain studies have generally focused on increased yield, which likely reflects lint yield being the cotton industry's top priority through history. For fiber quality, cultivar development programs have placed more emphasis on maintaining acceptable fiber quality levels while increasing yields, rather than directly increasing fiber quality *per se*. However, in recent years increasing fiber quality has become a higher priority for cultivar development programs because price premiums and discounts based on HVI fiber quality standards are being implemented.

Interestingly, improved lint yield and fiber quality potential during the 20th and 21st centuries has been attributed to inbreeding and phenotypic selection procedures based on capitalizing additive genetic effects (Campbell et al. 2008). Bridge and Meredith (1983) and Green and Culp (1990) recommended inter-mating adapted Upland genotypes and selecting novel allele combinations based on phenotypic selection to improve Upland cotton. This proposed method has been the predominant form of plant breeding over the last 50 years and is responsible for the documented genetic gains. However, this method has also led to reports documenting the decline of genetic diversity in the Upland cotton gene pool (Van Esbroeck, Bowman, Calhoun, and May 1998). Interestingly, the basis for the Pee Dee Upland and New Mexico Acala public cotton breeding programs, originally involved the introgression of genes from outside the Upland cotton gene pool (Bowman, May, and Calhoun 1996). Genotypes from *G. barbadense* and the diploid species *G. arboreum* and

G. thurberi via a triple hybrid species were hybridized to Upland genotypes and intermated to develop the basis of each breeding program, respectively. However, traditional cotton breeding approaches relying on hybridization and phenotypic selection have reported great difficulty introgressing stable and useful genes from inter-specific crosses within the primary tetraploid cotton gene pool (Saha, Jenkins, Wu, McCarty, Gutierrez, Percy, Cantrell, and Stelly 2006). Successful introgressive breeding strategies of the future will require approaches different than traditional, mainstream plant breeding (Paterson, Bowman, Brown, Chee, Gannaway, Gingle, May, and Smith 2004).

4 Molecular Genetics of Fiber Quality Properties

Cotton, like most other economically important crop species, has been extensively studied using molecular genetic tools. The objectives of most cotton genomics projects involve developing infrastructure such as DNA marker tools and genetic populations. This core infrastructure can then be used to dissect the quantitative variation of traits important to cotton production such as fiber productivity and quality, with the ultimate goal to apply marker-assisted selection (MAS) for cotton improvement.

As mentioned in the previous section, the main fiber properties that are highly correlated with spinning performance and end product quality, such as fiber length, strength, and fineness, are quantitatively inherited. As such, the segregating progeny resulting from specific crosses do not segregate into discrete classes like simply inherited traits. Instead, these traits show a continuous range of phenotypes in a segregating population as a result of the different number of favorable alleles that constitute a quantitative trait, each acting independently and with varying effects, present in the progenies. The individual factors that constitute a quantitative trait are called quantitative trait loci, or QTLs. Individual QTLs are difficult to study and to manipulate in classical breeding programs which rely solely on phenotypic selection because: (i) the traits are conditioned by the aggregate effects of other phenotypes, each of which are influenced by one or more genetic loci; (ii) the degree of phenotypic expression conferred by each genetic locus can vary depending on different environmental conditions; and (iii) individual genetic loci that associate with a quantitative phenotype often have genetic background effects due to interlocus interactions or epistasis. An understanding of quantitative variation at the molecular level could provide detailed descriptions of individual genes, and this is made possible by QTL mapping.

The principles underlying QTL mapping have been explained elsewhere (see Paterson, Lander, Hewitt, Peterson, Lincoln, and Tanksley 1988; Knapp and Bridges 1990; Tanksley 1993). There are two general requirements to perform QTL mapping: (i) a genetic map with sufficient polymorphic loci that cover the whole nuclear genome, and (ii) a discrete, genetic population such as an F₂,

backcross, or recombinant inbred population, that segregates for the quantitative trait. Basically, the analytical approach involves the detection of linkage disequilibrium via statistical association between alleles at a marker locus and alleles at the linked QTL. The presence of a QTL is declared when significant differences are observed between the means of phenotypic values of all individuals within a bi-parental population carrying different parental marker alleles at a genetic locus. Because QTL discovery is essentially a statistical process, a QTL is not described with the same precision as single locus. Instead, a QTL is defined as a segment of chromosome flanked by genetic markers affecting the quantitative trait of interest. Nevertheless, unlike classical quantitative genetic studies, which can only describe the aggregate effects of all the genes influencing a trait, QTL analysis allows the estimation of the number of genes, their location in the genome, and more importantly, the phenotypic and genetic effects of individual QTLs. Identification of individual QTLs could lead to improved selection efficacy in breeding programs, especially for traits with low to medium heritability or that are difficult, time consuming, or costly to measure.

4.1 Mapping QTL for Fiber Quality

The availability of computer software using algorithms for which the location and effects of individual QTLs can be more precisely resolved (Lander and Botstein 1989; Knapp and Bridges 1990; Knott and Haley 1992; Zeng 1994), and the recent development of saturated genetic maps that permit comprehensive analysis of the cotton genome (Rong, Pierce, Waghmare, Rogers et al. 2004; Nguyen, Giband, Brottier, Risterucci, and Lacape 2004; Cai, Guo, Wang, Han et al. 2007) have allowed the detection of many QTLs influencing fiber properties (discussed below). The number of QTLs conferring fiber quality that have been identified to date is listed in Tables 4 – 7, summarized below, and include 107 QTLs for fiber length; 13 QTLs for short fiber content; 83 QTLs for fiber strength; 33 QTLs for micronaire; and 112 QTLs for fiber fineness. DNA markers tightly linked to these QTLs could allow breeders to “design” a genetically superior individual based on the “best” DNA marker allele combinations to express a given phenotype. This process would provide cotton breeders with the ability to identify genotypes with the desired allele combinations from a large heterogeneous population.

The most efficient method for mapping QTLs in plants has been based on defined segregating populations such as F₂, recombinant inbred or backcross. These genetic populations are constructed by mating two carefully selected parents with divergent genetic backgrounds and contrasting differences for the trait of interest (e.g.- parents that will generate a large number of polymorphic loci that will segregate in the progeny and are fixed for the alternate alleles at the QTL). A fair number of QTLs have been mapped using

Table 4 Summary of QTL mapped for fiber length

Parents	No. QTL	Sub-genome			PVE (%)	Ref
		A	D	Unk		
CAMD-E x Sea Island Seaberry*	1		1		14.7	A
MD5678ne x Prema	2			2	11.5/44.6	B
Siv'on x F-177*	6			6	2.9–13.7	C
TM-1 x 3–79*	3	1	2		7.8–12.6	D
TM-1 x 3–79*	1		1		19.7	E
Siv'on x F-177*	6	5	1		2.9–13.7	F
Acala-44 x Pima S-7*	1	1			24.0	G
Tamcot 2111 x Pima S6*	28	13	14	1	6.0–30.0	H
TM-1 x 3–79*	1	1			5.2	I
Yumian 1 x T586	2	1	1		27.4/25.5	J
Handan208 x Pima90*	4	2	1	1	17.1–27.5	K
7235 x TM-1	5	1	4		7.6–14.1	L
HS427-10 x TM-1	3	2	1		8.6–20.2	L
PD6992 x SM3	3	1	2		5.5–18.4	L
Guazuncho 2 x VH8*	15	6	9		3.8–14.8	M
7235 x TM-1	4	1	2	1	6.4–9.9	N
TM-1 x 3–79*	5	2	3		5.7–9.0	O
ZMS12 x 8891	4	2	1	1	5.3–21.8	P
Handan208 x Pima90*	5	4	1		11.0–48.8	Q
7235 x TM-1	8	2	5	1	3.9–7.7	R
Total	107	45	49	13		
Percent		42	46	12		

References:

- | | | |
|--------------------------|----------------------|-----------------------------|
| A. Jiang et al. 1998 | G. Mei et al. 2004 | M. Lacape et al. 2005 |
| B. Ulloa & Meredith 2000 | H. Chee et al. 2005b | N. Shen et al. 2006 |
| C. Saranga et al. 2001 | I. Park et al. 2005 | O. Frelichowski et al. 2006 |
| D. Kohel et al. 2001 | J. Zhang et al. 2005 | P. Wang et al. 2006 |
| E. Ren et al. 2002 | K. Lin et al. 2005 | Q. He et al. 2007 |
| F. Paterson et al. 2003 | L. Shen et al. 2005 | R. Shen et al. 2007 |

*Interspecific cross

Table 5 Summary of QTL mapped for fiber uniformity and short fiber content

Parents	No.	Sub-genome			PVE (%)	Ref
	QTL	A	D	Unk		
Short Fiber Content						
Tamcot 2111 x Pima S6*	8	3	4	1	7.0–13.0	A
ZMS12 x 8891	5	0	3	2	5.8–10.6	B
Total	13	3	7	3		
Percent		23	54	13		
Fiber Uniformity						
Tamcot 2111 x Pima S6*	9	6	3	0	9.0–24.0	A
Percent		67	33	0		

References:

- | | |
|----------------------|---------------------|
| A. Chee et al. 2005b | B. Wang et al. 2006 |
|----------------------|---------------------|

*Interspecific cross

Table 6 Summary of QTL mapped for fiber strength

Parents	No. QTL	Sub-genome			PVE (%)	Ref
		A	D	Unk		
CAMD-E x Sea Island Seaberry*	3		3		9.7–13.3	A
MD5678ne x Prema	3			3	10.6–24.6	B
TM-1 x 3–79*	4	1	3		10.4–23.1	C
Siv'on x F-177*	21	7	14		2.4–17.4	D
TM-1 x 3–79*	2		2		5.4/6.3	E
7235 x TM-1	1	1			18.5–53.8	F
Yumian 1 x T586	3	2		1	11.8–20.2	G
Handan208 x Pima90*	2	1	1		21.1/21.6	H
Guazuncho 2 x VH8*	12	5	7		4.4–16.2	I
7235 x TM-1	2		2		7.0–18.2	J
HS427-10 x TM-1	5	3	2		9.2–12.0	J
PD6992 x SM3	3	2	1		7.0–7.2	J
7235 x TM-1	3		3		4.5–26.0	K
TM-1 x 3–79*	6	1	5		5.4–6.5	L
ZMS12 x 8891	4	1	2	1	4.6–9.4	M
7235 x TM-1	7	1	5	1	4.3–16.2	N
Handan208 x Pima90*	3	2	1		15.3–37.1	O
Total	84	27	51	6		
Percent		32	51	7		

References:

- | | | |
|--------------------------|-----------------------|-----------------------------|
| A. Jiang et al. 1998 | F. Park et al. 2005 | K. Shen et al. 2006 |
| B. Ulloa & Meredith 2000 | G. Zhang et al. 2005 | L. Frelichowski et al. 2006 |
| C. Kohel et al. 2001 | H. Lin et al. 2005 | M. Wang et al. 2006 |
| D. Paterson et al. 2003 | I. Lacape et al. 2005 | N. Shen et al. 2007 |
| E. Zhang et al. 2003 | J. Shen et al. 2005 | O. He et al. 2007 |

*Interspecific cross

intraspecific *G. hirsutum* populations. However, because of the low level of DNA polymorphism within the *G. hirsutum* gene pool (Brubaker and Wendel 1994; Pillay and Myers 1999; Van Becelaere, Lubbers, Paterson, and Chee 2004), most QTL mapping studies have been conducted using interspecific populations, especially from crosses between *G. hirsutum* and *G. barbadense*. In addition to providing the DNA-level polymorphism needed to expedite genetic map construction, these two cultivated species are each prized for somewhat different characteristics. Breeding of *G. hirsutum* has focused on maximum yield and broad adaptation, while breeding for *G. barbadense* has emphasized fiber quality. Consequently, *G. barbadense* has fiber quality much superior to that of the more widely-grown *G. hirsutum*. However, the narrow range of environmental adaptation of *G. barbadense*, limited to irrigated regions in arid zones of Western U.S., preclude its use in most of the US 'Cotton Belt.' Nevertheless, the unique fiber properties of *G. barbadense* make it an ideal candidate for providing new genetic variation to improve fiber quality in *G. hirsutum* and using DNA markers to mitigate the problems associated with such crosses.

Table 7 Summary of QTL mapped for fiber fineness/Micronaire

Parents	No.	Sub-genome			PVE (%)	Ref
	QTL	A	D	Unk		
Micronaire						
Tamcot 2111 x Pima S6*	9	1	8		9.0–25.0	A
7235 x TM-1	4		4		6.0–21.0	B
HS427-10 x TM-1	3	1	2		6.0–12.5	B
PD6992 x SM3	2	1	1		8.7/8.8	B
7235 and TM-1	2		1	1	11.1/16.1	C
Handan208 and Pima90*	7	3	4		16.2–28.9	D
Handan208 and Pima90*	8	3	5		9.7–21.5	E
ZMS12 and 8891	7	1	6		5.4–15.6	F
MD5678ne x Prema	4			4	6.2–21.7	G
TM-1 x 3–79*	6	4	2		16.7–43.9	H
Yumian 1 and T586	2	2			15.9/7.4	I
TM-1 and 3–79*	5	4	1		6.8–17.2	K
Siv'on x F-177*	25	10	15		2.2–30.3	L
Guazuncho 2 and VH8*	21	12	9		4.6–29.1	M
TM-1 and 3–79*	2	2			6.2/7.9	N
7235 and TM-1	5	1	3	1	4.5–10.9	O
Total	112	45	61	6		
Percent		40	55	5		
Fiber fineness						
Tamcot 2111 x Pima S6*	32	13	18	1	7.0–36.0	A
Acala-44 x Pima S-7*	1			1	43.2	J
Total	33	13	18	2		
Percent		39	55	6		

References:

A. Draye et al. 2005	F. Wang et al. 2006	K. Frelichowski et al. 2006
B. Shen et al. 2005	G. Ulloa & Meredith 2000	L. Paterson et al. 2003
C. Shen et al. 2006	H. Kohel et al. 2001	M. Lacape et al. 2005
D. Lin et al. 2005	I. Zhang et al. 2005	N. Park et al. 2005
E. He et al., 2007	J. Mei et al. 2004	O. Shen et al. 2007

*Interspecific cross

4.1.1 QTLs for Fiber Length

Fiber length is the most extensively studied fiber property using QTL mapping. Therefore, it is not surprising that more QTLs have now been identified for fiber length than any other fiber property as 107 QTLs were identified from 18 different QTL mapping studies by this review’s time of submission (Table 4). Collectively, these studies indicate that fiber length is governed by a relatively small number of QTLs with large effects. Many QTLs with predominantly minor effects account for a very large portion of the phenotypic variance. For example, only 8 of the 18 studies have identified one or more QTLs that explain more than 20% of the total phenotypic variance observed in the mapping population. Interestingly, more QTLs with large effects, those explaining 20% or more of the total phenotypic variance, were detected in intraspecific

populations (4/8 or 50%) than in interspecific populations (3/12 or 25%). The QTL with the largest effects detected in an intraspecific population was from the cross of MD5678ne and Acala Prema reported by Ulloa and Meredith (2000), which explained 44.6% of the phenotypic variation. These observations are unexpected because relatively little heterosis for fiber length exists in both modern and obsolete upland cultivars (Meredith 1990; Tang, Jenking, McCarty, and Watson 1993; Campbell et al. 2008), thus eliminating a large contribution via non-additive genetic effects. On the other hand, a greater number of QTLs were identified using populations derived from interspecific crosses, likely reflecting a greater number of loci segregating in these populations. However, more marker loci were segregating in the interspecific populations, which provide better genome coverage to allow for a more rigorous search for QTL. The large number of QTLs with small effects reflects the general complexity of this trait. These results are consistent with the consensus drawn from quantitative genetic variance analyses.

As mentioned previously, other length parameters such as fiber length uniformity and short fiber content are also important components of fiber quality. Analyses identified fewer QTLs for length uniformity and short fiber content suggesting the genetic component of these traits may be less than that of fiber length (Table 5). In addition to environmental influences and measurement errors, length uniformity and short fiber content may also have a greater sensitivity to variation in correlated fiber properties. For example, Behery (1993) showed that short fiber content could be an indirect consequence of fiber strength; cotton fibers with poor strength would have higher fiber breakage and therefore higher short fiber content, and lower fiber uniformity. However, Meredith (1996) studied the relationship between fiber strength and short fiber content, and determined that these two traits had a low correlation, which suggests that other genetic factors may be involved in causing short fiber content. Chee, Draye, Jiang, Decanini, et al. (2005a) reported the QTLs identified for both fiber length uniformity and short fiber content generally correspond with QTLs for fiber length. Thus, co-localization of QTLs may appear to account for the observed correlations among these traits.

From a crop improvement perspective, the large number of QTLs with small effects identified for fiber length reflects the general complexity of manipulating this quantitatively inherited trait. However, a majority of the QTLs identified appear to have predominantly additive effects. These results are consistent with the consensus drawn from quantitative genetic variance analyses in that additive variance is the most important genetic determinant of variation in fiber length (Wilson and Wilson 1975; Tang et al. 1993; May and Green 1994). Additive QTLs are thought to be most likely to continue to function as predicted when placed in other genetic backgrounds (Bernacchi, Beck-Bunn, Emmatty, Eshed, et al. 1998; Tanksley and Nelson 1996), although this remains to be tested in cotton.

4.1.2 QTLs for Fiber Strength

Eighty-four QTLs have now been identified for fiber strength based on 14 different QTL mapping studies (Table 6). The genetic basis for fiber strength appears to be similar to that of fiber length in that major QTLs exist in both inter- and intra-specific populations but a large portion of the phenotypic variation is conferred predominantly by minor QTLs. These results are not unexpected as several quantitative genetic studies have determined that the superior fiber strength in certain well-known cotton germplasm is conditioned by only a few major genes. For example, Meredith (1992) reported that the high fiber strength of the germplasm line MD51ne was conditioned by as few as two major genes. Since this line was derived from the germplasm line FTA 263, which has an origin tracing back to Beasley's triple hybrid *G. thurberi* X *G. arboreum* X *G. hirsutum* (Culp and Harrell 1980), it is thought that the fiber strength phenotype of MD51ne is likely to have derived from wild species gene introgression (Culp 1992).

Although QTLs with large effects are less likely to escape detection, the likelihood that the expression of fiber strength genes may be sensitive to environmental conditions presents a challenge in detecting these QTLs. The influence of G x E on fiber strength gene expression is corroborated by genotype based G x E analysis and the emerging QTL literature. In a QTL mapping experiment conducted in both well-watered and water-limited treatments tested in two different growing seasons, Paterson, Saranga, Menx, Jiang, and Wright (2003) reported that 21 fiber strength QTLs were mapped in a F2 and F2-3 population. However, 13 of the QTLs were significantly affected by environmental factors, including 6 QTLs showing significant effects in only one of the two growing seasons, and 7 QTLs were detected only under water-limited conditions. Zhang, Yuan, Yu, Guo, and Kohel (2003) conducted a similar QTL mapping study using an F2 and F2-3 population derived from the high strength parent 7235 and TM-1, and they observed a different outcome from those discussed above. They detected a major QTL for fiber strength located on chromosome 10, which explained 18.5% to 53.8% of the phenotypic variation. This QTL, which they suspect to have derived from Acala-type cotton, was stably expressed in four tested locations in China and the U.S.A. over two different growing seasons (Zhang et al. 2003). More recently, Shen, Guo, Lu, Zhu, Yuan, and Zhang (2007) tested a population of recombinant inbred lines derived from the F2s of Zhang et al. (2003) and detected the same fiber strength QTL as previously reported. Thus, it is possible to map QTLs for fiber strength that are stably expressed across different environments and over different generations.

4.1.3 QTLs for Fiber Fineness/Maturity

The main phenotype of interest is the physical property of cotton fiber related to the thickness or diameter of mature fibers. Direct determination of the physical

properties related to fiber diameter is now possible with single-fibered microscopic analyses, however the high cost in both time and labor have precluded its routine use in the textile industry (Bradow and Davidonis 2000). Several fiber-testing instruments have been developed that provide estimates of fiber diameter. As discussed in Section 2, micronaire has been the most widely used method of determining fiber diameter and is the standard technique utilized by the USDA cotton classing office in the U.S. Therefore, a majority of the QTL mapping studies conducted to date have used the micronaire measurement.

The number of QTLs mapped for fiber diameter based on micronaire is presented in Table 7. About 112 QTLs have now been identified for this phenotype from 16 different QTL mapping studies conducted via both inter- and intra-specific populations. The percentage of phenotypic variance explained by individual QTL ranged from 2.2% to 30.3%. Again, consistent with other fiber quality properties with complex inheritance, occasional QTLs with large effects have been detected, but a large portion of the phenotypic variation was conferred predominantly by loci with minor effects. Further, the number of QTLs detected was less than 10 in all except two studies, which were conducted using interspecific populations (21 and 25 QTLs reported by Lacape, Nguyen, Courtois, Belto, et al. 2005 and Paterson et al. 2003, respectively). The small number of QTLs detected for micronaire may suggest a high non-genetic influence for this measurement. May (2000) reported a similar conclusion after observing numerous quantitative genetic studies involving micronaire, which showed that genetic differences were smaller than non-genetic variation. Further, since micronaire is a measure of resistance to airflow of a constant weight of fibers, the micronaire value is confounded by both fineness and maturity (May, 2000). Low micronaire cotton could result from immature and/or very fine fibers, which cannot be distinguished.

Other more direct methods for measuring fiber diameter such as the AFIS measurement are now available. The AFIS instrument uses a light-attenuation signal to calculate the cross-sectional area of individual fibers, and therefore, offers a fiber diameter measurement not confounded with fiber maturity like that of the micronaire test (Bradow and Davidonis 2000; Hequet and Wyatt 2001). By using the AFIS data, Mei et al. (2004) reported the detection of a major QTL for fiber fineness explaining 43% of the phenotypic variation. In another study, Draye, Chee, Jiang, Decanini, et al. (2005) detected 32 QTLs with individual loci conferring 7% to 36% of the phenotypic variance. Interestingly, ten of the QTLs fell into the same region where Paterson et al. (2003) reported a QTL for micronaire, and 3 loci also corresponded to micronaire QTLs mapped in the same study (Draye et al. 2005). The detection of QTLs at corresponding chromosomal locations across various populations and using different methods of measurement further supports the validity of these fiber fineness QTLs. However, the low number of corresponding QTLs for micronaire and AFIS fiber fineness further supports the notion that the micronaire measurement can be confounded with other fiber properties such as maturity (Meredith 1994b). Therefore the degree of cell-wall thickening affected by fiber

maturity likely accounts for a portion of the unmapped genetic effects and explains the less accurate reflection of fiber thickness. Nonetheless, the cotton industry still relies extensively on micronaire readings to assess Fiber Fineness because of its speed and low cost.

4.2 *Meta-Analysis and Confirmation of QTLs*

Comparison of QTLs mapped in different populations by alignment to a common reference map may help to identify stably expressed QTLs across environments and genetic backgrounds. Comparison of QTL consistency across populations has been reported in numerous crop species (Groh, Gonzalez-de-leon, Khairallah, Jiang, et al. 1998; Melchinger, Utz, and Schon 1998; Monforte and Tanksley 2000). Population types and sample size, and the number and density of markers are important considerations in QTL mapping experiments. In addition, QTL detection is subject to experimental error that can be minimized by choosing the best statistical model that will optimize the power to detect phenotypic differences between two marker genotypes and setting appropriate false positive and false negative error rates. The use of different parameters listed above, along with the use of different parental combinations tested in different environments may account for the general lack of consistency in the number and location of QTLs reported in different studies.

In cotton, a number of studies have used a common set of DNA markers, particularly those based on restriction fragment length polymorphisms (RFLPs), in their genetic mapping experiments. These common genetic markers offer the opportunity to integrate the map information developed from different populations, including that of the location of QTLs for fiber quality properties. Two studies are particularly worth mentioning with respect to determining the consistency of fiber QTLs mapped in populations of different genetic backgrounds. First, Ulloa, Saha, Jenkins, Meredith, et al. (2005) reported on the development of the first joinmap of *G. hirsutum* using a core set of 111 RFLP markers that were mapped on four intraspecific cotton populations, including the placement of 145 QTLs, mostly for fiber properties, on the linkage map. Second, Rong, Wright, Saranga, May, et al. (2007) reported on the alignment of 432 fiber QTLs mapped in 10 interspecific *G. hirsutum* by *G. barbadense* populations onto a consensus map constructed based on the locus arrangement of the hypothetical ancestor that give rise to the diploid progenitors of tetraploid cotton. The meta-analysis from these two studies revealed several intriguing observations with regard to the genetic control of cotton fiber quality.

First, only about 10 to 20% of the QTLs detected in the one population were also detected in another population. This surprisingly low level of correspondence among fiber QTLs found in different populations suggests that lint fiber development may involve a complex gene network. For example, Rong et al.

(2007) compared the QTL maps detected in two interspecific F2 populations, which were developed using one common parent (Pima S7) and two closely-related genetic stocks containing different recessive mutations in fiber development: *n2*, a genetic stock with naked seed, and *im*, a genetic stock with immature fiber. Although the two populations were tested in the same environment, virtually no common QTLs were detected. This result shows that mutation in a fiber gene may impose particularly large perturbations, causing drastic regulatory changes in the expression of the phenotype.

The finding that fiber properties have a complex QTL landscape has three important implications. First, it suggests that QTL studies reported in the current literature are detecting only a small subset of the genes conferring fiber quality. Many mapping experiments attempting to detect fiber QTLs may have been conducted using insufficient DNA markers to cover the entire cotton genome or utilizing sample sizes that are too small to have the statistical power to effectively detect loci with small effects or to separate closely linked loci. This supposition is supported by Chee et al. (2005a,b) and Draye et al. (2005). These studies used a series of 24 advanced backcross families totaling 2,976 BC3F2 plants with 262 RFLP markers selected for even coverage of all 26 linkage groups of the cotton genome and reported a substantially larger number of QTLs for fiber length, fineness and elongation than prior QTL analyses conducted using F2 populations. Secondly, to gain a representative picture of the true complexity of fiber properties, one would need to study a broad sampling of germplasm in a wide range of environments. Because cotton improvement has relied on breeding for inbred lines by crossing closely related parents followed by backcrossing or direct reselection from existing cultivars (Bowman 2000), mutations that give rise to fiber genes with large effects are likely to have been genetically fixed within a germplasm group by selection. Thus, different germplasm groups would carry unique sets of genes that only will respond well to the specific environment in which the germplasm is adapted. Finally, from the crop improvement standpoint, it is highly advisable to validate a fiber QTL in the genetic background in which they are to be deployed prior to committing resources to making selections based on the favorable QTL allele. This issue will be revisited and expanded in the next section.

A second aspect of the meta-analysis reveals that the genes responsible for fiber quality appear to be non-randomly distributed across the cotton genome. Rong et al. (2007) showed that 47% of the QTL for fiber quality were located in 21% of the cotton genome. A similar observation was made by Ulloa et al. (2005) who noted that of the 47 linkage groups that were constructed in their joinmap, the linkage groups belonging to chromosome 3 and 26 were found to harbor nearly 50% of the QTLs for agronomic and fiber quality properties. Because these QTLs affecting different phenotypes were mapped in populations of diverging genetic background, the clustering tendency is not likely to be due to pleiotropy. Rong et al. (2007) speculated that the observed clustering of fiber QTLs may “represent groups of coordinately regulated genes and/or groups of

small gene families that have undergone proximal duplication followed by sub- or neo-functionalization”.

A third noteworthy observation emerging from the meta-analysis concerns the contribution of the A- and D-subgenome of allotetraploid cottons to the genetic variation of the fiber properties. Only the present day A-genome diploid species produces spinnable fiber. Therefore, the role of the D-genome in genetic determination of fiber quality in polyploid cotton has long been a question. The extent to which genetic variation in fiber quality properties is determined by loci on the D-subgenome was first provided by Jiang, Wright, El-Zik, and Paterson (1998) who showed that QTLs for fiber quality such as length, strength and fineness were associated with DNA markers mapped on the D-subgenome. In fact, a majority of the loci were found in the D-subgenome rather than the A-subgenome (Jiang et al. 1998). From an evolutionary genetic standpoint, this observation is fascinating because the genes responsible for the unique properties of lint fiber, the flat convoluted ribbon structure that permits the fiber to spun into yarn, likely first arose from natural mutations in the progenitor of the A genome diploid lineage prior to polyploid formation (Brubaker, Paterson and Wendel 1999). Evidence to support this hypothesis includes the fact that within the diploid *Gossypium* species, only the A genome species *G. arboreum* and *G. herbaceum* produce spinnable lint fibers. In addition, all allotetraploid *Gossypium* species have some form of convoluted seed hairs, but none of the New World D-genome species display this unique fiber property (Kohel, Quisenberry, Cartwright, and Yu 2000). Therefore, this character is purportedly attributed to their A-genome progenitor.

As shown in Tables 4 to 7, numerous QTL analyses have now confirmed the observation that the D-subgenome, from the ancestor that did not have any spinnable fiber, plays a large role in the genetic control of fiber growth and development in polyploid cotton (but see Ulloa et al. 2005). Collectively, these data demonstrate that D-subgenome genes have been recruited to contribute to the genetic control of fiber quality properties. Jiang et al. (1998) speculated that since the A-subgenome has a much longer history of selection for fiber development, many of the favorable alleles at major loci for fiber properties might have already been fixed as a result of natural selection. Presumably the formation of an elongated fiber promotes seed dispersal. On the other hand, loci for fiber development on the D-subgenome may not have come under heavy selection until after polyploid formation, and therefore, mutations that enhanced this trait may have become favorable only after polyploidization. Hence, when artificial selection was recently imposed via domestication and breeding, the D-subgenome may have more abundant allelic variation that confers a fitness advantage for fiber properties. Recruitment of D-subgenome loci may allow additional flexibility of polyploid cotton for artificial selection through breeding and thereby explains the superiority of the fiber properties of polyploid cottons relative to their cultivated A-genome diploid species.

5 Targetting QTLs For Improving Fiber Quality – Opportunity and Future Challenges

The use of DNA markers in mainstream cotton breeding has become routine in private cotton breeding programs. This is especially true in transgenic breeding where DNA markers are used in combination with backcrossing to select for multiple, homozygous transgene alleles to rapidly stack multiple transgenic traits into a single genotype. In a coordinated fashion, DNA markers are also used to select for background recurrent parent alleles in backcross progenies to speed up the process of re-capturing the recurrent parent genetic background. Thus, DNA markers allow cotton breeders to determine which progeny are to be retained or discarded for the next generation. According to US patent no. 6,376,753, the integration of DNA markers as a selection tool has been cited to significantly increase the efficiency and decrease the time a transgene is introgressed into an elite cultivar. Although DNA markers have not routinely been deployed in public improvement programs, the approach has successfully been utilized to develop germplasm with high levels of resistance to reniform nematodes (Robinson, Bell, Dighe, Menz, Nichols, and Stelly 2007). These examples are for simply inherited traits and in the case of transgenic breeding the actual transgene is used for direct selection, nonetheless they document the benefit that can be reaped by applying DNA markers in applied cotton breeding. Given that considerable information for markers that are associated with QTLs for fiber properties is now available, introgressing fiber quality traits is the most obvious application of DNA markers in molecular breeding programs.

With technological advancement in DNA genotyping, as RFLPs give way to simple sequence repeats (SSRs) and more recently to single nucleotide polymorphisms (SNPs), an enticing question is, can DNA markers be used to assist in improving quantitatively inherited traits such as fiber properties. Abundant fiber property QTLs have now been mapped and tightly linked DNA markers identified (see Section 4), yet the authors are not aware of any completed efforts where a fiber quality QTL has been successfully applied in MAS. As discussed in the previous section, a low level of correspondence among QTLs for fiber properties detected in different populations and testing environments highlights that each QTL should be subjected to further analysis if the goal is to establish a MAS program. In addition, the detection of co-locating QTLs detected in different genetic backgrounds and populations supports the notion that they occur in orthologous genomic regions. While this finding does increase the confidence level that a genomic region declared as a QTL is authentic, it does not prove that the co-localized QTLs are controlled by the same gene or closely linked genes. Therefore, it is critical that each locus be independently validated prior to being exploited by introgressing into elite germplasm.

A number of strategies have been developed to validate QTL position and effect. Two widely used approaches include measuring QTL effects in additional mapping populations, or backcrossing the QTL into one or more

genotypes to create sets of near-isogenic lines to estimate QTL effects (Anderson, Liu, and Cho 2007). Unfortunately, regardless of which QTL validation approaches are being utilized, the process is very time consuming and the outcome still may not adequately represent QTL effects in different genetic backgrounds. Therefore, it is of little surprise that while increasing numbers of fiber property QTLs continue to be discovered and reported each year, the efforts to validate these QTLs are not moving apace.

A small number of fiber property QTLs have now been validated for their authenticity and improvement value – these QTLs represent the best candidates for MAS. One such example is the QTL for fiber strength of the germplasm line 7235 identified by scientists from the Nanjing Agricultural University in China. 7235 has a complex pedigree including wild species such as *G. anomalum*, and *G. hirsutum* cultivars with high fiber strength, such as Acala 3080 and PD4381 (Qian, Huang, Peng, Zhou, et al. 1992). Zhang et al. (2003) conducted a bulk segregant analysis for fiber strength using F2 and F2:3 populations derived from a cross between 7235 and TM-1. A major fiber strength QTL with additive gene action, *QTL_{FS1}*, was detected on chromosome 10 by association with 8 DNA markers. Variance explained by this QTL ranged from 19 to 53% in three diverse environments in the U.S. (College Station, TX), and in China (Nanjing and Hainan). Interestingly, despite having wild species in its genetic background, the fiber strength QTL in 7235 was determined to have originated from Acala3080. In a subsequent study, Shen et al. (2005a,b) performed a comprehensive QTL analysis on the same F2 population reported Zhang et al. (2003) by increasing the marker density to 127 loci. The additional SSR markers allowed them to further narrow the fiber strength QTL to about 5 cM interval flanked by the markers BNL3474 and NAU1322. They later constructed a population of recombinant inbred lines by selfing the F2 progenies from the same 7235 by TM-1 cross (Shen et al. 2007). Again, the *QTL_{FS1}* locus was found to be stably expressed in the two test environments (Nanjing and Guanyun County, China) and over two years of field trials. The chromosomal location for this QTL was also revised to LGD8 due to a past error in assigning several SSR loci to a chromosome, and the QTL was renamed to *qFS-D8-1*. The phenotypic variance explained by this locus ranged from 4.3 to 16.2%, or about half of the average effect observed in the F2 and F2:3 generations (Zhang et al. 2003). Based on these results, Shen et al. (2007) reached the conclusion that *qFS-D8-1* is a stable QTL that is expressed across environments and generations, and suggested this QTL can be effectively used in MAS to improve fiber strength.

Identifying the map position of a QTL is only the first of many steps required for planning a MAS program. There are several challenges that need to be addressed if a breeding program desires to integrate DNA markers and apply MAS for fiber quality traits (Bohn, Groh, Khariallah, Hoisington, Utz, and Melchinger 2001; Melchinger et al 1998; Tanksley and Nelson 1996; Van Berloo and Stam 1999; Young, Schupp, and Keim 1999). Assuming that a fiber quality QTL position is correctly identified and that a tightly linked marker is available,

some of the other questions that need to be addressed include the following. (1) Will the QTL work in other genetic backgrounds? Since different cross combinations will have different sets of QTLs segregating, the effects of a target QTL may or may not be realized due to other genes segregating in the background, either acting epistatically with the target locus or losing other favorable alleles from the recurrent background. (2) Will the QTL have the same phenotypic effects in other genetic backgrounds? Most QTL mapping projects were performed by using population sizes (100–300 individuals) that were too small to accurately estimate individual QTL effects. The phenotypic effects predicted by a marker-QTL association may be overestimated in most cases (Beavis 1994; Bohn et al. 2001; Melchinger et al. 1998). (3) Is the QTL linked to other undesirable traits? Because most QTL mapping studies focus on one to a few target traits, the presence of undesirable genes linked to a QTL is often unknown. (4) Will the QTL-marker linkage relationship hold in different generations? Simulation studies showed that MAS is less efficient when compared to phenotypic selection if practiced over several successive generations, due to the higher rate of fixation of unfavorable alleles at QTLs with small effects in later generations, and to increased risk of recombination between markers and target QTLs (Hospital, Moreau, Lacoudre, Charcosset, and Gallais 1997). (5) How does the cost and efficiency of applying MAS compare to phenotypic selection alone? In the private sector, this question is certainly of utmost importance when considering implementing MAS in cultivar development. As rapid advances in DNA marker development and genotyping systems continue, evidence suggests that phenotyping and phenotypic selection may represent the largest cost factor in future breeding programs. Additional studies, such as Bernardo and Yu (2007), must be conducted to determine relative cost and genetic gain comparisons between phenotypic selection alone and some combination of MAS and phenotypic selection. The successful outcome of a MAS program hinges on a positive response to the questions above.

In summary, despite the fact that MAS has received considerable attention from breeders in the last decade, applying it is still rare. Many of our current obstacles in implementing MAS to improve fiber quality are due to the lack of complete understanding of the phenotypic effect of the individual QTLs. The current approaches to validate QTL position and effects are time consuming, expensive and may not adequately predict the effect of the QTL in our elite germplasm. Fortunately, with the rapid advancement in genomic tool development such as the 4,000 SSRs markers currently available in cotton, with more being developed each year, it seems certain that fine mapping of important QTLs for fiber properties is achievable in the near future although none have been reported in the literature. In addition to accurate mapping of the QTLs for fiber properties and development of DNA markers that are well-suited (closely linked and user-friendly) for MAS, fine mapping is also an essential step toward the final isolation of the QTL by map-based cloning (Zhang 2007). The isolated genes are the most desirable form of marker for MAS because there is no recombination between the gene and its derived marker. Given the progress

that we have achieved in understanding and locating QTLs for fiber properties in the past decade, we believe that molecular breeding has a bright future in cotton and MAS is expected to make an effective complement to phenotypic screening in the future genetic gain for fiber quality.

References

- American Society for Testing and Materials (1993) Standard test method for linear density and maturity index of cotton fibers (IIC-Shirley fineness/maturity tester). ASTM D3818–92.
- Anderson, J.A., Liu, S., and Cho S. (2007) Molecular breeding using a major QTL for fusarium head blight resistance in wheat. *Crop Sci.* 47(S3): 113–119.
- Backe, E.E. (1996) The importance of cotton fiber elongation on yarn quality and weaving performance. In C. Chewning, ed. *Proceedings of the 9th annual Engineered Fiber Selection System Conference*. Cotton Incorporated, Raleigh, NC.
- Baenziger, P.S., Russell, W.K., Graef, G.L., and Campbell, B.T. (2006) Improving Lives: 50 Years of Crop Breeding, Genetics, and Cytology (C-1). *Crop Sci* 46: 2230–2244.
- Bassett, D.M., and Hyer, A.H. (1985) Acala cotton in California: 60 years of varietal improvement. In *Proc. Beltwide Cotton Prod. Res. Conf.*, New Orleans, LA. Natl. Cotton Counc. Am., Memphis, TN. p. 76.
- Beavis, D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *Proceedings of the 49th annual corn and sorghum industry research conference*. Washington, DC: American Seed Trade Association, 250–266.
- Behery, H.M. (1993) Short fiber content and uniformity index in cotton. In: *International Cotton Advisory Committee Review Article No. 4*, CAB International, Wallingford, UK.
- Benzina, H., Hequet, E., Abidi, N., Gannaway, J., Drean, J.Y., and Harzallah, O. (2007) Using fiber elongation to improve genetic screening in cotton breeding programs. *Textile Research* 77: 770–778.
- Bernacchi, D., Beck-Bunn, T., Emmatty, D., Eshed, Y., Inai, S., Lopez, J., Petiard, V., Sayama, H., Uhlig, J., Zamir, D. and Tanksley, S.D. (1998) Advanced backcross QTL analysis of tomato. II. Evaluation of near isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor. Appl. Genet.* 97: 170–180.
- Bernardo, R. and Yu, J. (2007) Prospects for genome wide selection for quantitative traits in maize. *Crop Sci.* 2007 47: 1082–1090.
- Bohn, M., Groh, S., Khariallah, M.M., Hoisington, D.A., Utz, H.F., and Melchinger, A.E. (2001) Re-evaluation of the prospects of marker-assisted selection for improving insect resistance to *Diatraea* spp. in tropical maize by cross validation and independent validation. *Theor. Appl. Genet.* 103: 1059–1067.
- Bowman, D.T. (2000) Attributes of public and private cotton breeding germplasm. *J. Cotton Sci.* 4: 130–136.
- Bowman, D.T., May, O.L., and Calhoun, D.S. (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Sci.* 36: 577–581.
- Bradow, J. and Davidonis, G. (2000) Quantification of Fibre Quality and the Cotton Production-Processing Interface. *J. Cotton Sci.* 4: 34–64.
- Bridge, R.R., and Meredith, W.R. (1983) Comparative performance of obsolete and current cotton cultivars. *Crop Sci.* 23: 949–952.
- Bridge, R.R., Meredith, W.R., and Chism, J.F. (1971) Comparative performance of obsolete varieties and current varieties of upland cotton. *Crop Sci.* 11:29–32.
- Brubaker, C.L., Paterson, A.H., and Endel, J.F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42: 184–203.

- Brubaker, C.L., and Wendel, J.F. (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear Restriction-Fragment-Length-Poly-morphisms (RFLPs). *Am. J. Bot.* 81: 1309–1326.
- Cai, C.P., Guo, W.Z., Wang, C.B., Han, Z.G., Song, X.L., Wang, K., Niu, X.W., Wang, C., Lu, K.Y., Shi, B., and T.Z., Zhang. (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176: 527–541.
- Campbell, B.T., and Jones, M.A. (2005) Assessment of genotype \times environment interactions for yield and fiber quality in cotton performance trials. *Euphytica* 144: 69–78.
- Campbell, B.T., Bowman, D.T., and Weaver, D.B. (2008) Heterotic effects in top crosses of modern and obsolete cotton cultivars. *Crop Sci.* 48: 593–600.
- Cheatham, C.L., Jenkins, J.N., McCarty, J.C., Watson, C.E., and Wu, J. (2003) Genetic variances and combining ability of crosses of American cultivars, Australian cultivars, and wild cottons. *J. Cotton Sci.* 7: 16–22.
- Chee, P., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, R., Smith, C.W., and Paterson, A.H. (2005a) Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach. III. Fiber Length. *Theor. Appl. Genet.* 111: 772–781.
- Chee, P., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, R., Smith, C.W., and Paterson, A.H. (2005b) Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: I Fiber Elongation. *Theor. Appl. Genet.* 111: 757–763.
- Culp, T.W., and Green, C.C. (1992) Performance of obsolete and current cultivars and Pee Dee germplasm lines of cotton. *Crop Sci.* 32: 35–41.
- Culp, T.W., Harrell, D.C., and Kerr, T. (1979) Some genetic implications in the transfer of high strength genes to upland cotton. *Crop Sci.* 19: 481–484.
- Culp, T.W., and Harrell, D.C. (1980) Registration of medium staple cotton germplasm. *Crop Sci.* 20: 290.
- Deussen, H. (1992) Improved cotton fiber properties – The textile industry's key to success in global competition. In: C.R. Benedict and G.M. Jividen (Eds.), *Proceedings from Cotton Fiber Cellulose: Structure, Function and Utilization Conference*, Natl. Cotton Counc. Am., Memphis, TN. pp. 43–63.
- Draye, X., P. Chee, C. Jiang, L. Decanini, T. Delmonte, R. Bredhauer, C.W. Smith, and A.H. Paterson. (2005). Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach. II Fiber Fineness. *Theor. Appl. Genet.* 111: 764–771.
- Frelichowski, J.E., Palmer, M.B., Main, D., Tomkins, R., Cantrell, C. and Ulloa, M. (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. *Mol. Gen. Gen.* 275: 479–491.
- Green, C.C., and Culp, T.W. (1990) Simultaneous improvement of yield, fiber quality, and yarn strength in Upland cotton. *Crop Sci.* 30: 66–69.
- Groh, S., Gonzalez-de-leon, D., Khairallah, M.M., Jiang, C., Bergvinson, D., Bohn, M., Hoisingtio, D.A., and Melchinger, A.E. (1988) QTL mapping in tropical maize: III. Genomic regions for resistance to *Diatraea* spp. and associated traits in two RIL populations *Crop Sci.* 38: 1062–1072.
- He, D.H., Lin, Z.X., Zhang, X.L., Nie, Y.C., Guo, X.P., and Zhang, Y.X. (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* \times *G. barbadense*. *Euphytica* 153: 181–197.
- Hertel, K.L. (1940) A method of fibre-length analysis using the fibrograph. *Textile Research* 10: 510–525.
- Hertel, K.L. (1953) The Stelometer, it measures fiber strength and elongation. *Textile World* 103: 97–260.
- Hertel, K.L., and Craven, C.J. (1951) Cotton fineness and immaturity as measured by the Arealometer. *Textile Research Journal* 21: 765–774.

- Hequet, E., and Wyatt, B. (2001) Relationship among image analysis on cotton fiber cross sections, AFIS measurements and yarn quality. In: *Proceedings Beltwide Cotton Research Conference*, National Cotton Council of America, Memphis, TN. pp. 1294–1298.
- Hoskinson, P.E., and Stewart, J.M. (1977) Field performance of two obsolete cotton cultivars. In: *Proceedings Beltwide Cotton Research Conference*. Natl. Cotton Counc. Am., Memphis, TN. pp. 78–79.
- Hospital, F., Moreau, L. Lacoudre, F. Charcosset, A., and Gallais, A. (1997) More on the efficiency of marker-assisted selection. *Theor. Appl. Genet.* 95: 1181–1189.
- ICAC (International Cotton Advisory Committee). (2005) The Outlook for Cotton Supply in 2005/06. Washington, DC: Secretariat of the International Cotton Advisory Committee.
- Jiang C.X., Wright, R.J., El-Zik, K.M. and Paterson, A.H., (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. USA* 95: 4419–4424.
- Johnson, B. (1952) Use and application of fiber and spinning tests. *Natl. Cotton Counc. Am.*, Memphis, TN.
- Kloth, R.H. (1998) Analysis of commonality for traits of cotton fiber. *J. Cotton Sci.* 2: 17–22.
- Knapp, S.J., and Bridges, W.C. (1990) Using molecular markers to estimate quantitative trait locus parameters: Power and genetic variances for unreplicated and replicated progeny. *Genetics* 126: 769–777.
- Knott, S. A. and Haley, C. S. (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* 132: 1211–1222.
- Kohel, R.J., Quisenberry, J.E., Cartwright, G., and Yu, J. (2000) Linkage analysis of transgenes inserted into cotton via *Agrobacterium tumefaciens* transformation. *J. Cotton Sci.* 4: 65–69.
- Kohel, R.J., Yu, J., Park, Y.H., and Lazo, G. R. (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica* 121: 163–172.
- Lacape J.M., Nguyen, T.B., Courtois, B., Belot, J.L., Giband, M., Gurlot, J.P., Gawryziak, G., Roques, S., and Hau, B. (2005) QTL analysis of cotton fiber quality using multiple *G. hirsutum* x *G. barbadense* backcross generations. *Crop Sci.* 45: 123–140.
- Lander, E.S., and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., and Stewart, J. McD. (2005) Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. *Plant Breed.* 124: 180–187.
- May, L. (2000) Genetic variation in fiber quality. In A. S. Basra (Ed.), *Cotton fibers, development biology, quality improvement and textile processing*. Food Products Press, New York. pp. 183–230.
- May, O.L., and Green, C.C. (1994) Genetic variation for fiber traits in elite Pee Dee cotton populations. *Crop Sci.* 34: 684–690.
- May, O.L., and Taylor, R.A. (1998) Breeding cottons with higher yarn tenacity. *Textile Res.* 68: 302–307.
- McCarty, J.C., Jenkins, J.N., and Wu, J. (2004) Primitive Accession Derived Germplasm by Cultivar Crosses as Sources for Cotton Improvement: I. Phenotypic Values and Variance Components. *Crop Sci.* 44: 1226–1230.
- Mei, M., Syed, N.H., Gao, W., Thaxton, P.M., Smith, C.W., Stelly, D.M., and Chen, Z.J. (2004) Genetic mapping and QTL analysis of fiber related traits in cotton (*Gossypium*). *Theor. Appl. Genet.* 108: 280–291.
- Melchinger, A.E., Utz, H.F., and Schon, C.C. (1998) Quantitative trait locus mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimating effects. *Genetics* 39: 546–557.
- Meredith, W.R. (1984a) Genotype x environment interactions. In: Kohel J.K. and Lewis, C.F., (Eds.), *Cotton*. Vol. 24. Amer. Soc. Agr., Madison. pp. 138–141.

- Meredith WR (1994b) Genetics and management factors influencing textile fiber quality. In: Chewing C (Ed.) Proc 7th Ann Cotton Incorporated Engineered Fiber Selection System Res. Forum, Cotton Incorporated, Raleigh, N.C., pp 256–261.
- Meredith, W.R. Jr. (1991) Associations of maturity and perimeter with micronaire. In J. M. Brown, ed. *Proceedings of the Beltwide Cotton Conference*. Natl. Cotton Counc. Am., Memphis, TN. pp. 569.
- Meredith, W.R. Jr. (2005) Minimum Number of Genes Controlling Cotton Fiber Strength in a Backcross Population. *Crop Sci.* 45: 1114–1119.
- Meredith, W.R., and Brown, J.S. (1998) Heterosis and combining ability of cottons originating from different regions of the United States. *J. Cotton Sci.* 2: 77–84.
- Meredith, W.R. Jr. (1990) Yield and fiber quality potential for second-generation cotton hybrids. *Crop Sci.* 30: 1045–1048.
- Meredith, W.R. Jr. (1992) Improving fiber strength through genetics and breeding. In C.R. Benedict (ed.) *Proc. Cotton Fiber Cellulose: Structure Function and Utilization Conf.*, Savannah, GA. Natl. Cotton Council Am., Memphis, TN. pp. 289–302.
- Meredith, W.R. Jr. (1996) Agronomic Factors and Yield Variability. *Proc. Natl. Cotton Counc. Beltwide Cotton Conferences*. Nashville, TN, pp. 180–184.
- Monforte, A.J., and Tanksley, S.D. (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome1 affecting fruit characteristics and agronomic traits: Breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theor. Appl. Genet.* 100: 471–479.
- Nguyen, T.B., Giband, M., Brottier, P., Risterucci, A.M., and Lacape, J.M. (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109: 167–175.
- Park, Y.H., Alabady, M.S., Ulloa, M., Sickler, B., Wilkins, T.A., Yu, J., Stelly, D.M., Kohel, R.J., El-Shihy, O.M., and Cantrell, R.G. (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred (RIL) cotton population. *Mol. Genet. Gen.* 274: 428–441.
- Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S. Lincoln, S.E, and Tanksley, S.D. (1998) Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726.
- Paterson, A.H., Boman, R.K., Brown, S.M., Chee, P.W., Gannaway, J.R., Gingle, A.R., May, O.L., and Smith, C.W. (2004) Reducing the genetic vulnerability of cotton. *Crop Sci.* 44: 1900–1901.
- Paterson, A.H., Saranga, Y., Menz, M., Jiang, C.X., and Wright, R.J. (2003) QTL analysis of genotype x environment interaction affecting cotton fiber quality. *Theor. Appl. Genet.* 106: 384–396.
- Pillay, M., and Myers, G.O. (1999) Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Sci.* 39: 1881–1887.
- Pressley, E.H. (1942) A cotton fiber strength tester. *American Society for Testing and Materials Bulletin* 118: 13–17.
- Qian, S.Y., Huang, J.Q., Peng, Y.J., Zhou, B.L., Ying, M.C., Shen, D.Z., Liu, G.L., Hu, T.X., Xu, Y.J., Gu, L.M., Ni, W.C., and Chen, S. (1992) Studies on the hybrid of *G. hirsutum* L. and *G. anomalum* Wawr. & Peyr. and application in breeding (in Chinese). *Sci. Agric. Sinica* 25: 44–51.
- Ren, H.L., Guo, W.Z., and Zhang, T.Z. (2002) Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line. *Acta Botanica Sinica* 44: 815–820.
- Robinson, A.F., Bell, A.A., Dighe, N.D., Menz, M.A., Nichols, R.L., and Stelly, D.M. (2007) Introgression of resistance to nematode *Rotylenchulus reniformis* into Upland Cotton (*Gossypium hirsutum*) from *Gossypium longicalyx*. *Crop Sci.* 47: 1865–1877.
- Rong, J., Wright, R.J., Saranga, Y., May, O.L., Wilkins, T.A., Draye, X., Waghmare, V.N., Feltus, F.A., Chee, P.W., Pierce, G.J., and Paterson, A.H. (2007) Meta-analysis of

- polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics* 176: 2577–2588.
- Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C., Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schulze, S.R., Trolinder, N.L., Wendel, J.F., Wilkins, T.A., Williams-Coplin, D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L., and Paterson, A.H. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166: 389–417.
- Saha, S., Jenkins, J.N., Wu, J., McCarty, J.C., Gutierrez, O.A., Percy, R.G., Cantrell, R.G. and Stelly, D.M. (2006) Effects of chromosome-specific introgression in Upland cotton on fiber and agronomic traits. *Genetics* 172: 1927–1938.
- Saranga, Y., Menz, M., Jiang, C., Wright, R., Yakir, D., and Paterson, A. H. (2001) Genomic dissection of genotype x environment adaptation conferring adaptation of cotton to arid conditions. *Genome Res.* 11: 1988–1995.
- Shen, X.L., Guo, W.Z., Zhu, X.F., Yuan, Y.L., Yu, J.Z., Kohel, R.J., and Zhang, T.Z. (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol. Breed.* 15: 169–181.
- Shen, X., Guo, W., Lu, Q., Zhu, X., Yuan, Y., and Zhang, T. (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. *Euphytica* 155: 371–380.
- Shen, X.L., Zhang, T.Z., Guo, W.Z., Zhu, X.F., and Zhang, X.Y. (2006) Mapping Fiber and Yield QTLs with Main, Epistatic, and QTL by Environment Interaction Effects in Recombinant Inbred Lines of Upland Cotton. *Crop Sci.* 46: 61–66.
- Smith, C.W., and Coyle, G.G. (1997) Combining ability for within-boll yield components in cotton, *Gossypium hirsutum* L. *Crop Sci.* 37: 1118–1122.
- Tang, B., Jenkins, J.N., McCarty, J.C., and Watson, C.E. (1993) F₂ hybrids of host plant germplasm and cotton cultivars. II. Heterosis and combining ability for fiber properties. *Crop Sci.* 33: 706–710.
- Tang, B., J.N. Jenkins, McCarty, J.C., and Watson, C.E. (1997) Evaluation of genetic variances, heritability, and correlations for yield and fiber properties among cotton F₂ hybrids. *Euphytica* 91: 315–322.
- Tanksley, S.D., and Nelson, J.C. (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92: 191–203.
- Tanksley, S.D. (1993) Mapping polygenes. *Annu. Rev. Genet.* 27: 205–233.
- Ulloa, M., and Meredith, W.R.Jr. (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *J. Cotton Sci.* 4: 161–170.
- Ulloa, M., Saha, S., Jenkins, J.N., Meredith, W.R. Jr., McCarty, J.C. Jr., and Stelly, D.M. (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) Joinmap. *J. Hered.* 96: 132–144.
- Van Becelaere, G., Lubbers, E.L., Paterson, A.H., and Chee, P.W. (2005) Pedigree vs. RFLP based genetic similarity estimates in cotton. *Crop Sci.* 45: 2281–2287.
- Van Berloo, R.V., and Stam, P. (1999) Comparison between marker-assisted selection and phenotypical selection in a set of *Arabidopsis thaliana* recombinant inbred lines. *Theor. Appl. Genet.* 98: 113–118.
- Van Esbroeck, G., Bowman, D.T., Calhoun, D.S., and May, O.L. (1998) Changes in the genetic diversity of cotton in the USA from 1970 to 1995. *Crop Sci.* 38: 33–37.
- Wang, B.H., Guo, W.Z., Zhu, X.F., Wu, Y.T., Huang, N.T., and Zhang, T.Z. (2006) QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton. *Euphytica* 152: 367–378.
- Wendel, J.F., and Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78: 139–186.

- Wilson, .FD., and Wilson, R.L. (1975) Breeding potentials of noncultivated cottons. I. Some agronomic and fiber properties of selected parents and their F1 hybrids. *Crop Sci.* 15: 763–766.
- Young, W.P., Schupp, J.M., and Keim, P. (1999) DNA methylation and AFLP marker distribution in the soybean genome. *Theor. Appl. Genet.* 99: 785–790.
- Yuan, Y.L., Zhang, T.Z., Guo, W.Z., Pan, J.J., and Kohel, R.J. (2005) Diallel analysis of superior fiber quality properties in selected upland cottons. *Acta Genetica Sinica* 32: 79–85.
- Zhang, H.B. (2007) Map-based cloning of genes and QTLs. In: *Plant Molecular Mapping and Breeding*. C. Kole and A. Abbott (eds.). Springer (in press)
- Zhang, J.F., Lu, Y., Adragna, H., and Hughs, E. (2005) Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. *Crop Sci.* 45: 2363–2373.
- Zhang, T.Z., Yuan Y.L., Yu J., Guo, W.Z., and Kohel, R.J., (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. *Theor. Appl. Genet.* 106: 262–268.

Bridging Classical and Molecular Genetics of Cotton Disease Resistance

Robert J. Wright, Chen Niu, and Bay Nguyen

Abstract Understanding the series of events that causes a symptomatic disease response brings to light the relationship between a pathogen and its host. More than 50 bacteria, fungi, virus, and nematodes have been implicated in cotton diseases that annually account for an estimated \$897 million in lost revenue to the U.S. crop. Advances in cotton genetics have led to the development of several technologies that enable the detailed examination of the cotton genome for genes that function in response to pathogen infection. There is an opportunity to make direct and meaningful comparisons from data generated by forward and reverse genetics. Comparing QTL information with reverse genetics (functional genomics) offers a powerful approach to identify and characterize the key pathways and the genetics involved in disease. Cotton is an interesting taxon in which to study disease because its evolution includes the divergence of resistance (R-genes) through the radiation of species, as well as multiple genetic bottlenecks including polyploid formation and the domestication of a small subset of *Gossypium* species.

1 Introduction

Understanding the series of events that cause a symptomatic disease response will bring to light the relationship between a pathogen and its host. In cotton (*Gossypium* spp.), disease has been a tertiary problem that affects plant productivity. Both insect pests and abiotic stress limit production to a greater extent than disease. However, losses due to disease reduce US cotton production by an estimated 11% annually (Blasingame 2006). Considering annual U.S. production of 22 million bales sold at \$0.70 per lb, the annual loss in revenue to producers is approximately \$897 million due to disease.

R.J. Wright (✉)

Department of Plant and Soil Sciences, Texas Tech University
e-mail: robert.wright@ttu.edu

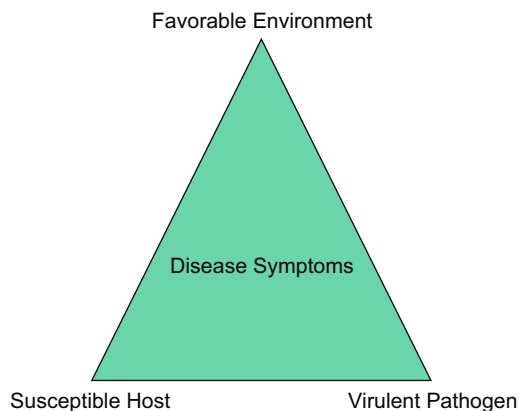


Fig. 1 Disease Triangle

Many discoveries have enabled present-day scientists to begin addressing questions regarding the mechanisms of disease. Phytopathology is an interesting science because two organisms must come together in a favorable environment to elicit a symptomatic response. This relationship is often represented by the Disease Triangle (Fig. 1). Disease symptoms will only occur when a susceptible host and virulent pathogen come together under conditions that favor infection and pathogen development. Disrupting this pivotal relationship is an important strategy to control disease.

The question of why some species cause disease (pathogenic) and others do not has been a primary focus of researchers. The pathogenic organism(s) of most cotton diseases have been identified, characterized, and in a few cases placed in well defined subgroups (pathotypes or races). Pathogen diversity and the genes (virulent and avirulent) that cause disease are focal areas of research. Likewise, the sources of host genes that confer resistance to the pathogen have been a key target. Cotton is an interesting taxon in which to study disease because its evolution includes the divergence of resistance (R-genes) through the radiation of species, as well as multiple genetic bottlenecks including polyploid formation and the domestication of a small subset of *Gossypium* species. These evolutionary events have shaped the genetics of resistance in cotton and resulted in distinct areas of coevolution with the pathogen. This chapter will briefly examine the current state of genetic research regarding cotton and the pathogens that cause disease.

2 Disease Losses and Management

More than 50 bacteria, fungi, virus, and nematodes have been implicated in cotton disease (Table 1). Several excellent review articles (Innes 1983; Delannoy, Lyon, Marmey, Jalloul, Daniel, Montillet, Essenberg and Nicole 2005; Davis,

Table 1 Cotton Diseases

Pathogen	Disease
<i>Agrobacterium tumefaciens</i>	Crown Gall
<i>Alternaria alternata</i>	Alternaria Leaf Spot
<i>Alternaria gossypina</i>	Alternaria Leaf Spot
<i>Alternaria macrospora</i>	Alternaria Leaf Spot
<i>Ascochyta gossypii</i>	Ascochyta Blight
<i>Ashbya gossypii</i>	Internal boll infection
<i>Aspergillus flavus</i>	Aflatoxin
<i>Aspergillus flavus</i>	Aspergillus boll rot
<i>Aspergillus nomius</i>	Aflatoxin
<i>Aspergillus parasiticus</i>	Aflatoxin
<i>Belonolaimus</i> Spp	Sting Nematodes
<i>Cercospora gossypina</i>	Cercospora Leaf Spot
<i>Colletotrichum</i> spp	Anthrachnose boll rot
<i>Fusarium oxysporum</i> f. sp. <i>vasinfectum</i>	Fusarium Wilt
<i>Fusarium</i> spp.	Seedling Disease
<i>Fusarium</i> spp.	Fusarium Boll Rot
<i>Glomerella gossypii</i>	Anthrachnose
<i>Hoplolaimus columbus</i>	Lance Nematodes
<i>Lasiodiplodia theobromae</i>	Lasiodiplodia(Diplodia)boll rot
<i>Leveillula taurica</i>	Powdery Mildew
<i>Macrophomina phaseolina</i>	Charcoal Rot
<i>Meloidogyne incognita</i>	Root-Knot Nematodes
<i>Myrothecium roridum</i>	Myrothecium Leaf Spot
<i>Nematospora coryli</i>	Internal boll infection
<i>Nigrospora oryzae</i>	Nigrospora Lint Rot
<i>Pantoea agglomerans</i>	Pantoea Boll Rot
<i>Phakopsora gossypii</i>	Tropical Cotton Rust
<i>Phomopsis</i> spp.	Phomopsis Leaf Spot
<i>Phymatotrichopsis omnivora</i>	Phymatotrichum Root Rot
<i>Phytophthora</i> spp.	Phytophthora boll rot
<i>Puccinia cacabata</i>	Southwestern Cotton Rust
<i>Puccinia schedonnardi</i>	Tumblegrass Rust
<i>Pythium</i> spp.	Seedling Disease
<i>Ramularia gossypii</i>	Areolate or False Mildew
<i>Rhizoctonia solani</i>	Seedling Disease
<i>Rhizoctonia solani</i>	Rhizoctonia Leaf Spot
<i>Rotylenchulus reniformis</i>	Reniform Nematodes
<i>Salmonia malachrae</i>	Powdery Mildew
<i>Sclerotium rolfsii</i>	Southern Blight
<i>Stemphylium solani</i>	Stemphylium Leaf Spot
<i>Thielaviopsis basicola</i>	Black Root Rot
<i>Verticillium dahliae</i>	Verticillium Wilt
Viral	African cotton mosaic
Viral	Cotton anthocyanosis
Viral	Cotton Blue disease
Viral	Cotton Leaf Curl

Table 1 (continued)

Pathogen	Disease
Viral	Cotton Yellow Vein Disease
Viral	Flavescence
Viral	Flower virescence
Viral	Psyllosis
Viral	Small Leaf
<i>Xanthomonas campestris</i> pv. <i>malvacearum</i>	Bacterial Blight
<i>Xanthomonas campestris</i> pv. <i>malvacearum</i>	Xanthomonas boll rot
?	Bronze Wilt

Colyer, Rothrock and Kochman 2006), books (Hillocks 1992; Kirkpatrick and Rothrock 2001), and state extension websites address the symptomatology and epidemiology of each disease. Disease is considered a problem in all cotton growing regions especially in counties with historical trends (i.e. blue disease in Brazil and leaf curl disease in Pakistan) or when seasonal climatic conditions elevate potential risks (i.e. early season cool temperatures). Minimizing yield losses due to insects and available water (drought) are the major management considerations in almost all cotton producing regions. However, diseases can cause substantial yield loss and reduced fiber quality if management decisions fail to identify or consider potential problems and risks. In the USA alone, annual yield loss due to diseases average 2,668,262 bales since 1995 (Table 2) (Blasingame 2006).

Considering that the average value of the US cotton crop is about \$7.7 billion (\$0.70 per lb) per year, disease accounts for an \$897 million loss per year. Yield losses in other countries usually exceed those seen in the USA (Kirkpatrick and Rothrock 2001). Nematodes are the most detrimental to yield, followed by seedling disease and boll rot (Table 2).

The management (control) of disease has relied on chemical, agronomic, and genetic strategies. The use of high quality seed and planting when soil temperatures favor vigorous plant development will help establish a healthy plant stand (Kirkpatrick and Rothrock 2001). The stringent management of weed and insect pests as well as rotation with alternative crops has reduced losses to several key pathogens. The use of fungicide treated seed or fungicide treatments during planting has been effective in reducing losses to seedling pathogens such as *Pythium* spp., *Rhizoctonia solani*, *Thielaviopsis basicola*, *Fusarium* spp., and others.

Cotton producing regions have relied on intrinsic resistance as a means to reduce the incidence of disease epidemics, although this management practice has only been effective for a few cotton disease pathogens. This is primarily because few genotypes resistant to most diseases have been identified in cultivated *G. hirsutum* and *G. barbadense* cotton. The evolution of cotton through the radiation of species, polyploidization, and domestication has unintentionally reduced the genetic variability for resistance. When resistant germplasm is available, breeding of resistant cultivars has been an effective strategy to reduce disease (Bird 1982; Ulloa, Hutmacher, Davis, Wright, Percy and Marsh 2006).

Table 2 Estimated Cotton Disease Losses 1995–2005

Disease	Mean	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Fusarium Wilt	^a 0.39	0.42	0.31	0.32	0.45	0.41	0.50	0.40	0.38	0.45	0.33	0.33
Verticillium Wilt	98,573 0.78	73,748 1.22	84,862 1.29	77,760 1.37	87,137 0.77	103,717 0.75	122,585 0.62	121,944 0.41	93,382 0.53	128,765 0.42	109,040 0.62	81,358 0.54
Bacterial Blight	162,717 0.13	254,246 0.10	241,984 0.10	287,831 0.11	158,529 0.13	153,304 0.25	134,877 0.16	53,065 0.20	111,866 0.26	71,687 0.05	187,948 0.02	134,553 0.01
Phymatotriconium	76,119 0.40	41,162 0.51	42,616 0.54	62,657 0.57	34,983 0.31	168,824 0.31	98,841 0.31	142,456 0.31	240,413 0.25	1,337 0.38	890 0.44	3,129 0.50
Root Rot												
Seedling Disease	258,904 2.93	222,212 3.82	167,020 2.75	261,100 3.54	113,828 2.78	180,846 2.78	163,909 2.43	158,329 2.71	186,716 3.35	241,707 4.02	517,727 2.18	634,553 1.88
Ascochyta Blight	626,031 0.13	828,199 0.09	549,628 0.04	634,861 0.47	440,771 0.13	551,946 0.09	589,655 0.04	704,639 0.13	721,387 0.20	854,777 0.06	540,327 0.10	470,154 0.09
Boll Rots	12,915 2.38	8,712 2.45	5,374 2.73	4,416 1.58	9,204 2.35	9,519 2.14	3,769 1.66	24,219 3.82	30,587 3.65	6,909 2.06	16,671 2.25	22,686 1.49
Nematode Spp.	478,179 4.10	462,359 3.54	555,675 3.42	307,503 3.71	339,866 3.53	348,882 4.24	282,256 4.39	1,082,587 4.13	575,919 4.22	428,828 4.54	503,728 4.68	372,368 4.71
Leaf Spots and Others	879,404 0.41	778,703 0.68	763,066 0.69	762,520 0.64	625,200 0.23	757,215 0.22	788,991 0.35	979,357 0.51	950,610 0.44	1,059,848 0.20	1,029,656 0.33	1,178,280 0.24
Total Percent	75,419 11.65	139,401 12.83	95,060 11.86	101,249 12.30	48,646 10.68	44,949 11.20	50,693 10.46	110,677 12.62	104,110 13.28	62,157 12.18	61,658 10.95	59,610 9.79
Bales Lost	2,668,262	2,808,742	2,505,287	2,557,304	1,858,163	2,319,202	2,235,576	3,377,273	3,014,990	2,856,015	2,967,645	2,956,691

^aTable entries are % loss and bales lost

The development of cultivars that show little or no disease symptoms in the presence of a pathogen is a major goal of plant breeders. Generally, single genes with large effects have been used to confer resistance in crop plants. Single-gene resistance usually provides immunity or a high level of resistance that is often effective against only a subset of a specific pathogen population (i.e. pathotype or race specific). Changes in the virulence of the pathogen can overcome the effects of this resistance, placing all genotypes with such “vertical” resistance at risk of infection by a new virulent pathotype or race. However, exceptions exist in which single gene resistance has remained stable for considerable periods of time (Nelson 1978). The development of stable resistance to pathogens is a matter of serious concern in plant breeding and has inspired debates among peers as to which gene deployment strategy most benefits the agricultural community, while minimizing any additional risk of disease epidemics.

It is apparent the potential sources of resistance can be found in exotic *Gossypium* species. Identifying and deploying genes from feral *Gossypium* species has been used to improved resistance to the Bacterial Blight pathogen *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) and recently to the reniform nematode, *Rotylenchulus reniformis* (Knight 1953, 1963; Innes 1983; Robinson, Bell, Dighe, Menz, Nichols and Stelly 2007). In these examples resistance was identified in diploid cotton and transferred into tetraploid types.

3 Disease Resistance in *Gossypium*

A mounting record of research regarding the discovery and characterization of resistant *Gossypium* germplasm includes of book citations, conference abstracts, extension newsletters, cultivar/germplasm releases, and peer reviewed scientific journal articles. Regrettably, this history is usually limited to a broad characterization of resistance among the *Gossypium* species or to a few key sources of resistant germplasm. A strong foundation of well characterized R-genes and inheritance studies is limited to only a few diseases, most notably the bacterial blight research conducted by various scientists (Delannoy et al. 2005). “Cotton Disease” (Hillocks 1992) provides an excellent record of research that predates the recent advent of molecular and genomic research into host-pathogen interactions in *Gossypium*.

The phenotypic complexity of many cotton diseases has perplexed and deterred many scientists. Generally, the single gene race specific resistance model does not apply to most cotton diseases. In these gene-for-gene models, the phenotypic response is usually discrete (immunity or a high level of resistance) that is often effective against only a subset of a specific pathogen population (i.e. pathotype or race specific). The inheritance of such resistant types is usually qualitative and can be explained by simple Mendelian models. The majority of cotton diseases are better explained by quantitative models with a continuous degree of phenotypic response to infection, and complex

inheritance. This may be partly explained by imprecise screening assays, measures of resistance that are influenced by multiple genetic (i.e. yield) and environmental factors, and undetected or un-characterized pathogen variability (mixed strains). It is clear that research and breeding activities cannot rely solely on natural infection to assess resistant and susceptible disease reactions because the incidence of disease symptoms is inconsistent and will result in the potential mis-classification of susceptible types (i.e. disease escapes). Interactions between the A and D subgenomes of tetraploid cotton may have contributed to this complexity. Tetraploid cotton stems from a recent polyploidy event about 1–2 million years ago. This event may have altered the relationship among cotton and its pathogens. Since this is a recent evolutionary event it is likely that a sufficient level of subgenomic divergence and pathogen co-evolution has yet to occur. This is somewhat supported by a different and less complex inheritance in diploid cotton. Molecular mapping has revealed that the genetic basis of cotton-pathogen interactions is more complex than classical data had suggested, and that the A and D-subgenomes have made very different contributions to this co-evolution (Wright, Thaxton, El-Zik and Paterson 1998).

The fragmentary data notwithstanding, research clearly suggests that R-gene evolution does not always parallel speciation and that genes (alleles) have been fixed, lost, or altered subsequent to *Gossypium* speciation events. The implications of this for breeding and germplasm enhancement are evident by the diversity of resistant and susceptible disease reactions within the *Gossypium* genus and by the deployment of genes from exotic sources (Knight 1953, 1963; Robinson et al. 2007).

Advances in cotton genetics have led to the development of several technologies that enable the detailed examination of the cotton genome for genes that function in response to pathogen infection. More than 185,000 ESTs (Udall, Swanson, Haller, Rapp, Sparks, Hatfield, Yu, Wu, Dowd, Arpat, Sickler, Wilkins, Guo, Chen, Scheffler, Taliercio, Turley, McFadden, Payton, Klueva, Allen, Zhang, Haigler, Wilkerson, Suo, Schulze, Pierce, Essenberg, Kim, Llewellyn, Dennis, Kudrna, Wing, Paterson, Soderlund and Wendel 2006) and 400 mapped quantitative trait loci (QTL) (Rong, Feltus, Waghmare, Pierce, Chee, Draye, Saranga, Wright, Wilkins, May, Smith, Gannaway, Wendel and Paterson 2007) are available for examination, however a disproportionately small number represent tools and information developed to study disease. A total of 3879 ESTs are available from libraries related to *Xanthomonas* and *Fusarium* infections (Patil, Pierce, Phillips, Venters and Essenberg 2005; Udall et al. 2006) (www.agcol.arizona.edu/pave/cotton). Thirty-six QTL or markers explain the variation of resistant and susceptible reactions to the pathogens that cause Bacterial Blight, Leaf Curl Virus, Verticillium Wilt, and nematodes (Wright et al., 1998; Aslam, Jiang, Wright and Paterson 1999; Rungis, Llewellyn, Dennis and Lyon 2002; Bolek, El-Zik, Pepper, Bell, Magill, Thaxton and Reddy 2005; Shen, Becelaere, Kumar, Davis, May and Chee 2006; Niu, Hinchliffe, Cantrell, Wang, Roberts and Zhang 2007; Yang, Guo, Li, Gao, Lin and Zhang 2008). The molecular dissection of cotton disease resistance has

provided some interesting and unexpected discoveries and continues to gain in popularity as new disease problems emerge (i.e. *Fusarium* Wilt epidemic in Australia) and cotton production expands into countries with unique disease problems.

4 Advances using Molecular Genetics

QTL mapping alone does not provide knowledge regarding the mechanisms and pathways involved in disease resistance. There is an opportunity to make direct and meaningful comparisons from data generated by forward and reverse genetics. Comparing QTL information with reverse genetics (functional genomics) offers a powerful approach to identify and characterize the key pathways and the genetics involved in disease. Recently, a community-wide effort produced Expressed Sequence Tags (ESTs) from cDNAs derived from tissues subjected to infection by *Xcm* and *Fusarium oxysporum* (Patil et al. 2005; Udall et al. 2006).

Pathogen virulence and population dynamics are key aspects of disease. Molecular genetics has enabled scientists to characterize races and vegetative compatibility groups in several pathogenic species of cotton. These analytic tools have the capacity to accelerate the identification of genes and gene products directly involved in disease. The assessment of a plant's response to pathogen infection is without doubt a key step when examining the consequences of host-pest interactions.

4.1 Bacterial Blight

Analysis of the subgenomic (A_t vs. D_t) distribution of genes conferring resistance to the bacterial blight pathogen, *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) (formerly designated *X. campestris* pv. *malvacearum*), in tetraploid AD-genome cottons provided an interesting system for studying the impact of allopolyploid formation on host-pathogen interactions (Wright et al. 1998). Among seven resistance genes derived from tetraploid cottons, six (86%) mapped to D-subgenome chromosomes. This suggests that the D-subgenome of tetraploid cotton has a higher propensity to give rise to new R-gene alleles and may indicate that polyploid formation offers novel avenues for phenotypic response to selection. Possible explanations of this bias include biogeographic factors, differences in evolutionary rates between subgenomes, gene conversion or other intergenomic exchanges that escaped detection by genetic mapping, or other factors (Wright et al. 1998). Genetic mapping revealed the chromosomal locations of the B_2 , B_3 , b_6 , and B_{12} genes that confer resistance to *Xcm* (Table 3). The complexity of the cotton-*Xcm* relationship is reflected in the discovery of both "horizontal" and "vertical" resistance components. In most cases, genetic mapping

Table 3 Summary of putative QTLs that confer resistance to *Xcm*

Population	Race ^a	Probable gene identity ^b	Nearest		LOD	% Variation
			DNA marker	Chr. or LG		
Empire B2	2 (4)	<i>B</i> ₂	<i>G1219</i>	D08	103.12	98.0
Empire B3	2 (4)	<i>B</i> ₃	<i>pGH510a</i>	D08	23.21	88.2
	18 (7)	<i>B</i> ₃	<i>pGH510a</i>	D08	10.56	53.4
Empire B2b6	2 (4)	<i>B</i> ₂	<i>G1219</i>	D08	53.36	92.2
	18 (7)	<i>Qb</i> _{6a}	<i>A1666</i>	D02	3.32	26.9
	18 (7)	(<i>Qb</i> _{6b})	<i>P12-20a</i>	5	3.07	22.4
	18 (7)	<i>Qb</i> _{6c}	<i>pAR827</i>	20	3.53	19.4
	18 (7)	(<i>Qb</i> _{6d})	<i>pAR723</i>	14	3.01	16.3
S295	18 (2, 4, 7)	<i>B</i> ₁₂	<i>pAR043</i>	14	50.46	94.2

^aParentheses indicate *Xcm* race phenotype which were highly correlated with the *Xcm* race phenotype used to calculate biometrical parameters.

^bParentheses indicate that the QTL had an effect opposite to that expected from the parental phenotype.

corroborated classical predictions regarding the number of genes, and gene action conferring resistance to various *Xcm* races. Only the “*b*₆ phenotype” conferring resistance to *Xcm* Races 7 and 18 exhibited numerous inconsistencies. The complexity of the *b*₆ phenotype, involving 4 genes (*Qb*_{6a}, *Qb*_{6b}, *Qb*_{6c}, and *Qb*_{6d}) with very different dosage effects and derived from different parents was an unexpected discovery that demonstrates a horizontal component of bacterial blight resistance. The renaming of several linkage groups (Rong, Abbey, Bowers, Brubaker, Chang, Chee, Delmonte, Ding, Garza, Marler, Park, Pierce, Rainey, Rastogi, Schultze, Trolinder, Wendel, Wilkins, Williams-Coplin, Wing, Wright, Zhao, Zhu and Paterson 2004) now places *B*₂ and *B*₃ loci on LGD08 (formerly designated Chromosome 20), *Qb*_{6a} on LGD02 (formerly designated LGU01), *Qb*_{6b} on Chromosome 5, *Qb*_{6c} on Chromosome 20 (formerly designated LGD04), *Qb*_{6d} and *B*₁₂ on Chromosome 14.

The genetic basis of the *B*_{6m}, *B*₆, or *b*₆ bacterial blight resistance gene resistance has been a subject of confusion among scientists and cotton breeders. The *B*_{6m} annotation has been historically used to describe an R-gene discovered in the A-genome species *G. arboreum* and introgressed into tetraploid cotton (*G. barbadense*) (Knight 1953). Knight reported that *B*_{6m} intensified resistance of *B*₂ but did not have any effect alone. Saunders and Innes (1963) further characterized *B*₆ as a gene that confers partial resistance. Others have used *B*₆ to describe a dominant R-gene. Wright et al. (1998) used genetic mapping to determine the chromosomal locations and subgenomic distributions of the *B*₂ and *B*₆ genes in a population (Empire B2b6) segregating for both genes. Genetic mapping corroborated classical predictions regarding *B*₂; however, molecular analyses of reaction to *Xcm* Races 2 or 4 in the Empire B2b6 population did not support the modified expression of the *B*₂ locus by *b*₆ that results in increased plant resistance. Segregation ratios in this population did not fit a two locus

(dominant/recessive) model. The QTL likelihood map supports a single genetic locus (B_2) explaining 92.2% of the phenotypic variation to *Xcm* Races 2 and 4. The complexity of the b_6 phenotype, involving 4 genes with very different dosage effects and derived from different parents, presumably explains the deviation from a simple segregation model that was observed. Based on classical evidence that the true b_6 allele is recessive and derived from an A-genome diploid (hence not likely to be on a D-genome chromosome), it was putatively linked to Qb_{6a} , on linkage group U01 of uncertain subgenomic origin. Subsequent mapping has now placed this gene on LGD02 and thus excluded it as Knight's original b_6 R-gene (Rong et al. 2004; Wright unpublished).

The B_{12} locus has been implicated as the contributing source of *Xcm* resistance in several Australian cotton cultivars (Rungis et al. 2002). The source of *Xcm* resistance in Australian cultivars was Tamcot germplasm (Tamcot SP37, Tamcot SP37H, and Tamcot SP23) developed by Texas A&M University (Luckett, Thomson and Reid 1986). The B_2 , B_3 , B_7 and B_{Sm} resistance genes were deployed in these cultivars (El-Zik and Bird 1970; Bird 1976). However, segregation analysis of resistant and susceptible progeny indicated that a single locus conferred Race 18 resistance in several Australian cultivars, notably Sicala V-2 and Siokra V-16. Molecular marker analysis revealed that Race 18 resistance was explained solely by the B_{12} locus in cultivar CS50 (Rungis et al. 2002). There is some evidence that resistance in S295 (Follin, Girardot, Mangano and Benitez 1988; Wright et al. 1998) and CS50 may be conferred by different alleles or closely linked loci. The source of CS50 resistance came from the line 101-102B which is susceptible to the highly virulent HVI isolates, unlike the resistant S295 cultivar. S295 is also a cultivar that has not been utilized by the Australian breeding program and therefore is not a contributing source of resistance in the Australian germplasm.

Bacterial blight is one of two diseases, the other being Fusarium wilt, for which ESTs have been developed in response to pathogen infection. More than 2000 clones (cDNAs) isolated from inoculated and non-inoculated leaf tissue was used to examine gene expression in Upland cotton following pathogen infection (Patil et al. 2005). Gene expression of inoculated and non-inoculated plants at 8, 14, 20, 30, 45, and 60 hpi (hours post inoculation) was used to screen for genes up-regulated in response to *Xcm* infection. Most genes (98%) were significantly up-regulated during the time-course and 63% were associated with defense response genes based on sequence homology. These libraries have been included in a community wide EST effort to assemble all *Gossypium* ESTs (Udall et al. 2006) using the Program for Assembling and Viewing ESTs (PAVE) (www.agcol.arizona.edu/pave/cotton). The mining of these ESTs for simple sequence repeats (SSRs) has revealed a few informative DNA marker loci that are available (Taliercio, Allen, Essenberg, Klueva, Nguyen, Patil, Payton, Millena, Phillips, Pierce, Scheffler, Turley, Wang, Zhang and Scheffler 2006).

A growing body of evidence suggests that all cotton bacterial blight R-genes react with the *avrBs3* family of avirulence genes (Yang and Gabriel 1995; De Feyter, McFadden and Dennis 1998; and Delannoy et al. 2005). This

observation suggests that the diversity among the R-genes is likely negligible and that once an R-gene is cloned and sequenced a candidate gene approach will quickly lead to the isolation of the remaining R-genes. Members of the *avrBs3* family contain 14 to 23 tandem repeats of a 102 bp leucine-rich sequence. The organization and number of repeats is believed to provide specificity that can elicit an incompatible response in the host (Leach and White 1997). Functional nuclear localization signals (NLSs) in the *avrBs3* gene suggest that avirulence proteins function inside plant cells (Yang and Gabriel 1995). Each gene is between 1.5 and 4.3 kb in size, and duplicated within the *Xanthomonas* genome. Twenty functional and nonfunctional *avrBs3* homologs have been isolated from *Xcm* strains (Gabriel 1999; Delannoy et al. 2005). The *XcmH* (moderately virulent) strain has ten members of the *avrBs3* gene family (De Feyter and Gabriel 1991; Yang, Yuan and Gabriel 1996) while the *XcmN* (highly virulent) strain contains fewer than 5.

4.2 Black Root Rot

Black root rot (BRR), incited by the soilborne pathogen *Thielaviopsis basicola* is an interesting system in which to study the evolutionary differences leading to R-gene formation and divergence in cotton (*Gossypium* spp). Resistant genotypes have not been available for the breeding of superior germplasm. This is attributed to a lack of resistance found in cultivated tetraploid cotton (*G. hirsutum* and *G. barbadense*); however, resistant types have been identified in the tetraploid progenitor species, *G. herbaceum* (Wheeler, Gannaway and Keating 1999). Approximately 1,000 tetraploid and diploid accessions were examined of which an A-genome diploid cotton species, *G. arboreum* (var. 'PI1415'), had reduced root necrosis when challenged with 337 to 550 chlamydospores/cm³ soil from *T. basicola* compared with a susceptible check. However, PI 1415 had significantly higher root necrosis ratings when challenged with 500 to 1,000 chlamydospores/cm³ soil compared with *G. herbaceum* var. 'A20' (Wheeler and Gannaway, 2007).

Genetic mapping has been used to detect the chromosomal locations of QTLs that confer resistance to the BRR pathogen. A population of 129 F₂ individuals (*G. herbaceum* x *G. arboreum*) and F_{2:3} progeny families were examined (Niu, Lister, Nguyen, Wheeler and Wright 2008). Genetic components that control resistant and susceptible reactions could be explained by three QTLs (Table 4). A region on LG.A5 (*BRR5.1*) explained 5.8 - 19.1% (LOD 2.74 - 5.75) of the phenotypic variation in three data sets. In each case, a likelihood peak was observed in the interval between markers *BNLI693* and *NAU1072*. The *G. herbaceum* allele improved resistance in each data set. Linkage Group 9 contained a second QTL (*BRR9.1*) that explained 8.2 - 10.3% (LOD 2.86 - 3.94) of the phenotypic variation in the F₂-adjusted and F_{2:3} data sets. The likelihood peak for each mapped within the interval delineated by markers *BNL2690* and *NAU0921*. The *G. herbaceum* allele increased resistance to *T. basicola*. A third

Table 4 Summary of putative QTLs that confer resistance to *Thielaviopsis basicola*

Population	QTL	L.G.	Interval ^a	Position ^b	LOD	% Var
F ₂	<i>BRR5.1</i>	A5	<i>BNL1693-NAU1072</i>	6.01	2.74	5.77
F _{2:3}	<i>BRR5.1</i>	A5	<i>BNL1693-NAU1072</i>	5.01	5.75	19.10
F ₂ -adjusted	<i>BRR5.1</i>	A5	<i>BNL1693-NAU1072</i>	6.11	3.93	7.22
F _{2:3}	<i>BRR9.1</i>	A9	<i>BNL2690-NAU0921</i>	2.01	2.86	10.31
F ₂ -adjusted	<i>BRR9.1</i>	A9	<i>BNL2690-NAU0921</i>	3.01	3.94	8.24
F _{2:3}	<i>BRR13.1</i>	A13	<i>BNL3442-BNL1034</i>	2.01	3.02	8.49

^aMarkers flanking the QTL likelihood peak.

^bPosition of the QTL likelihood peak (centi-Morgan from top).

QTL (*BRR13.1*) explained 8.49% of phenotypic variation among the F_{2:3} families. The interval between markers *BNL3442* and *BNL1034*, on LG.A13, contained the likelihood peak (LOD 3.02). The *G. herbaceum* allele contributed to the increased BRR resistance. The three QTLs collectively explained 32.7% of the phenotypic variation in reaction to *T. basicola* among F_{2:3} families.

Comparative analysis was conducted on significant QTL regions to deduce the cotton–*Arabidopsis* synteny relationship and examine the correspondence between BRR QTLs and *Arabidopsis* pathogen defense genes. Each QTL region was aligned with the corresponding cotton consensus map (Rong et al. 2005; <http://chibba.agtec.uga.edu/cgi-bin/cmap/viewer>) based on conserved marker loci to locate BRR QTLs on the consensus map (Niu et al. 2008). Markers that flanked the 99% likelihood interval (2-LOD) of each QTL were used in this step. Comparative analysis linked to synteny-based information revealed several regions of correspondence between BRR QTLs and *Arabidopsis* pathogen defense response genes. The inferred position of 56 *Arabidopsis* genes, including four defense and nine stress response genes, could be placed within a BRR QTL region. The CMap display tool (<http://chibba.agtec.uga.edu/cmap/>) is a very good resource to make comparisons between an individual map and the Cotton Consensus Map (Rong et al., 2005) which has linked QTL knowledge and cotton-*Arabidopsis* synteny (Rong et al., 2007). Although the position of all *Arabidopsis* genes could not be inferred in cotton, some defense and stress response genes could be located within BRR QTL intervals. This synteny based comparison may reveal underlying cotton defense response genes in cotton and an evolutionary link to a common ancestor.

4.3 *Fusarium Wilt*

Fusarium oxysporum f. sp. *vasinfectum* (Fov) Atk. W.C. Snyder & H.N. Hans is a worldwide soil fungal pathogen that caused Fusarium wilt in all cotton producing regions. Symptoms can appear at any stage of crop development and are compounded by pathogen virulence, host susceptibility, climatic conditions, soil type and fertility, and nematode populations (Davis et al 2006).

Until recently, the *Fov*-nematode complex was considered a key aspect leading to the severity of disease symptoms (Hillocks 1983; Hillocks and Kibani 2002). However, the virulent Australian *Fov* strain will cause disease in the absence of nematodes (Davis, Moore, and Kochman, 1996).

More than eight *Fov* races or biotypes have been characterized based on their specific virulence to several host species. The characterization of races has been called unorthodox and confusing because alfalfa, soybean, and tobacco have been used to differentiate certain *Fov* races (based on pathogenicity) which are otherwise identical among cotton genotypes. Pathogenicity testing on cotton for Races 1, 2, and 6 are indistinguishable as well as Races 3 and 5 (Davis et al. 2006). Several scientists have called for and are working toward a classification of races founded on their pathogenicity on cotton (Assigbetse, Fernandez, Dubois and Geiger 1994; Davis et al., 2006) and genetic diversity among strains (Bridge, Ismail and Rutherford 1993; Kim, Hutmacher and Davis 2005). The bridge between the classical and molecular characterization of *Fov* strains has recently been demonstrated. The genetic diversity among 28 isolates from a worldwide collection was examined by comparing the partial DNA sequences of the translation elongation factor (*EF-1 α*), phosphate permease (*PHO*), nitrate reductase (*NIR*), and the mitochondrial small subunit (mtSSU) (Skovgaard, Nirenberg, O'Donnell and Rosendahl 2001). Four distinct lineages were revealed that corresponded with differences in virulence and the geographic origin of each isolate. A second study involved examination of 30 isolates including two highly virulent strains from Australia by comparing the partial DNA sequences of the *EF-1 α* , *PHO*, and beta-tubulin (*BT*) genes (Kim et al. 2005). Five distinct lineages were identified, included the four initially described and a fifth that represented the Australian strains. Restriction fragment analysis of ribosomal and mitochondrial DNA (rDNA and mtDNA) was less definitive in the separation of isolates based on their pathogenicity (Fernandez, Assigbetse, Dubois and Geiger 1994). The selective amplification and restriction digestion of the intergenic spacer (IGS) region of nuclear rDNA, however, did reveal nine distinct haplotypes which were predictive of pathogenicity on cotton (Kim et al. 2005).

Although estimated disease losses approach 100,000 bales (\$34 million) annually in the United States, breeding for *Fov* resistant cultivars has not been a high priority compared to other diseases. The urgency to initiate research and breeding activities has changed due to the Australian *Fov* epidemic (Davis et al. 1996) and discovery of Race 4 in California (Kim et al. 2005). The Australian epidemic quickly spread via field run-off water and contaminated soil on equipment. The epidemic has been linked to two new pathogenic *Fov* strains that do not require nematodes for infection. An effort to identify resistant cotton genotypes was initiated leading to the discovery of *G. sturtianum* accessions (Gos-5275) with enhanced resistance compared to the most resistant *G. hirsutum* cultivar (McFadden, Beasley, and Brubaker 2004). Interestingly, susceptible *G. sturtianum* accessions were transgressive to the most susceptible *G. hirsutum* cultivar. These results suggest substantial genetic diversity for resistance within the *G. sturtianum* germplasm pool. This, however, may also be a consequence of

the diploid nature of *G. sturtianum*. Because nematodes have such a pivotal role in incidence and severity of Fusarium wilt, genetic improvements that increase nematode resistance in cotton have also seen reduced losses due to Fusarium Wilt. An effort to examine the inheritance of resistant and susceptible phenotypes when challenged with *Fov* has been initiated by USDA-ARS. Their goal is to identify DNA markers closely linked (associated) to gene(s) that confer resistance to Races 1 and 4. A recent report outlining progress can be accessed at www.ars.usda.gov/research/projects/projects.htm?ACCN_NO=409856&fy=2007.

4.4 Leaf Curl Virus

The cotton leaf curl virus (*CLCuV*) is a disease of cotton in Pakistan and the northern region of India. The symptomatic response to infection appears as upward curling of leaves and thickening of leaf veins (more pronounced on the underside). In extreme cases, the formation of a cupshaped or leaf laminar outgrowth called “enations” appears on the underside of the leaf. The *CLCuV* disease is transmitted by the whitefly, *Bemisia tabaci* (Genn) and it has many alternate hosts among cultivated and wild *Malvaceae* (Hameed, Khalid, Ehsan-ul-Haq and Hashrni 1993). The disease was recently shown to be caused by multiple monopartite *Begomoviruses* containing a single stranded DNA satellite molecule termed CLCuV DNA- β (Mansoor, Briddon, Bull, Bedford, Bashir, Hussain, Saeed, Zafar, Malik, Fauquet and Markham 2003).

Genetic mapping has been used to examine the inheritance of resistant and susceptible disease reactions in an F_2 cross between susceptible *G. barbadense* L. (Giza-45) and resistant *G. hirsutum* L. (Reba P-288) parents. Individual plant reactions to *CLCuV* were clearly bimodal (223 resistant: 62 susceptible) and did not deviate significantly from a single-gene model ($p > 0.2$) (Aslam, Jiang, Wright and Paterson 1999). These observations were consistent with prior research suggesting that a single dominant gene in *G. hirsutum* confers resistance to the disease (Ali 1997). A region on Chromosome 4 near the DNA markers *A1215*, *A1826*, and *pGH318* accounted for resistant and susceptible phenotypic reactions to *CLCuV*. Two DNA marker loci, detected by probes *A1215* and *A1826*, essentially co-segregate with the locus (0.0 and 0.1 cM). A third marker locus (*pGH318*) is 11.6 cM from the resistance locus. Since the publication of this research two separate research groups have published findings that suggest the inheritance is controlled by three loci (Rahman, Hussain, Malik and Zafar 2005; Ahuja, Monga, Dhayal 2007).

4.5 Verticillium Wilt

Detectable in most soils, the fungus *Verticillium dahliae* Kleb is a global problem for cotton producers and a focal point of recent research activities. The

phenotypic control of resistant and susceptible disease reactions has been characterized using simple (Barrow, 1970a, b; Pan, Zhang, Kuai, Guo and Wang 1994; Mert, Kurt, Gencer, Akiscan, Boyaci and Tok 2005) and complex (Verhalen, Brinkerhoff, Fun and Morrison 1971; Roberts and Staten 1972; Devey and Roose; 1987) genetic models. Recent studies using molecular marker technologies support an inheritance model involving several loci that individually explain less than 25% of the total phenotypic variation of resistant and susceptible individuals (Gao, Nie, and Zhang 2003; Du, Du and Ma 2004; Bolek et al., 2005; Wang, Zhang, He, Lin, Nie, Li and Chen 2005; Zhen, Wang, Ma, Zhang and Wang 2006; Yang et al., 2008). Measurable phenotypes (i.e. plant height, numbers of leaves, numbers of nodes, dry matter weights, etc.), skewed by environmental and unrelated genetic factors, loosely reflect disease reactions and likely contribute to the genetic complexity and low heritability estimates. However, this is beginning to change as scientists design experiments to minimize undesirable variability and maximize meaningful information. Fundamental information about the genetic control of *Verticillium* wilt resistance is accumulating, toward practical application in breeding.

Susceptible disease responses (symptomatology) when challenged with *Verticillium dahliae* are designated as defoliating (P-1) and non-defoliating (P-2) (Bell, 1994; Kirkpatrick and Rothrock 2001). There are four vegetative compatibility groups (VCG) within *V. dahliae* and each group has several subgroups. Isolates which fall into VCG 1A are known as defoliating whereas VCG 1B, 2A, 2B, 3, 4A, 4A/B, and 4B are non-defoliating (Kirkpatrick and Rothrock 2001).

Verticillium wilt has received much attention at the level of QTL mapping. Bolek et al. (2005) examined the number of healthy leaves, number of nodes, leaf weight, stem weight, and total shoot weight to assess resistant and susceptible disease reactions using Bulk Segregant Analysis of an F₂ population (*G. barbadense* x *G. hirsutum*). Resistance to a defoliating pathotype (V76) was associated with two QTLs on Chromosome 11 (Bolek et al. 2005) (Table 5). A QTL between markers *CM12* and *STS1* explained variation (16.5–29.5%) in

Table 5 Summary of putative QTLs that confer resistance to *Verticillium dahliae*

Population (Citation)	Trait ^a	QTL	L.G.	Interval ^b	LR ^c	% Var
F ₂ (Bolek et al, 2005)	Leaf	<i>LG-1</i>	11	<i>CM12-STS1</i>	12.78	17.0
	Node	<i>LG-1</i>	11	<i>CM12-STS1</i>	15.85	19.1
	Leaf Wt	<i>LG-1</i>	11	<i>CM12-CM29</i>	11.03	16.5
	Stem Wt	<i>LG-1</i>	11	<i>CM12-STS1</i>	18.10	29.5
	Shoot Wt	<i>LG-1</i>	11	<i>CM12-CM29</i>	14.32	21.9
	Leaf	<i>LG-2</i>	11	<i>NEJ6-BNL3147-2</i>	12.10	31.1
	Leaf Wt	<i>LG-2</i>	11	<i>NEJ6-BNL3147-2</i>	13.29	33.2

^aNumber of healthy leaves (Leaf); number of nodes (Node); leaf weight (Leaf Wt); stem weight (Stem Wt); total shoot weight (Shoot Wt)

^bMarkers flanking the QTL likelihood peak.

^cSignificant ($P < 0.05$) likelihood ratios (LR) based on 1000 permutation tests. LR of 9.20 = LOD of 2.

all traits measured (Table 5). The second QTL between markers NEJ6 and *BNL3147-2* explained variation for the number of healthy leaves (31.1%) and leaf weight (33.2%). In all cases the allele from the resistant parent, Pima S-7 (*G. barbadense*), conferred improved tolerance.

A recent study examined resistant and susceptible individuals in two related lineages (F_2 and BC_1S_2 families) using three separate *V. dahliae* isolates at two developmental stages (Yang et al., 2008). This study was innovative because it examined two different defoliating and a single non-defoliating *V. dahliae* isolate(s) independently on 96 BC_1S_2 families. This enabled the scientists to map putative resistance QTLs that explained the phenotypic variation caused by each isolate. A total of 20 QTLs explained resistant and susceptible phenotypes when challenged with these isolates (Table 6). There was no correspondence between QTLs mapped in the F_2 and BC_1S_2 populations when challenged with the non-defoliating BP2 isolate. This obscures a definitive comparison regarding the Chromosome 5 QTL which may confer resistance to both defoliating and non-defoliating isolates. Several common QTLs were mapped among the defoliating isolates that suggest a similar genetic control of resistance.

Yang et al. (2008) identified some congruence with two prior studies. A single SSR locus (BNL3255), putatively mapped to Chromosome 5 (Zhen et al. 2006)

Table 6 Summary of putative QTLs that confer resistance to *Verticillium dahliae*

Population (Citation)	Isolate	QTL	L.G.	Interval ^a	LOD	% Var
F_2 (Yang et al, 2008)	BP2	<i>qVL-A5-1F₂</i>	A5	JESPR65b-Y20c	2.35	8.0
	BP2	<i>qVL-A5-2F₂</i>	A5	NAU2029-NAU4106	3.31	12.3
	BP2	<i>qVV-A5-1F₂</i>	A5	NAU1406-NAU2014b	4.40	17.1
	BP2	<i>qVL-A7-1F₂</i>	A7	NAU474b-NAU1048	2.10	8.1
	BP2	<i>qVV-A7-1F₂</i>	A7	NAU474b-NAU1048	4.08	14.0
	BP2	<i>qVL-A8-1F₂</i>	A8	NAU3201-NAU2665	2.50	8.4
	BP2	<i>qVV-A9-1F₂</i>	A9	BNL3582-NAU1360	2.62	7.7
BC_1S_2 (Yang et al, 2008)	VD8	<i>qVL-A5-1BC₁S₂VD8</i>	A5	NAU5273-NAU569b	2.57	10.5
	VD8	<i>qVV-A5-1BC₁S₂VD8</i>	A5	NAU5273-NAU569b	4.16	15.4
	592	<i>qVL-A5-1BC₁S₂592</i>	A5	NAU3036-NAU2121	2.12	9.9
	592	<i>qVL-A5-2BC₁S₂592</i>	A5	NAU3607-NAU1065a	2.11	11.0
	BP2	<i>qVV-A8-1BC₁S₂BP2</i>	A8	NAU920-NAU3964	4.48	16.1
	VD8	<i>qVL-A8-1BC₁S₂VD8</i>	A8	JESPR232-NAU3201	4.06	15.0
	BP2	<i>qVL-D4-1BC₁S₂BP2</i>	D4	NAU3437-NAU3392	5.49	20.3
	BP2	<i>qVV-D4-1BC₁S₂BP2</i>	D4	NAU3437-NAU3392	2.33	7.8
	VD8	<i>qVV-D5-1BC₁S₂VD8</i>	D5	NAU1042-NAU828b	3.03	10.6
	592	<i>qVL-D5-1BC₁S₂592</i>	D5	NAU2513-BNL1878	2.44	14.1
	592	<i>qVV-D5-1BC₁S₂592</i>	D5	BNL2656-BNL1671	2.04	8.5
	VD8	<i>qVV-D11-1BC₁S₂VD8</i>	D11	NAU643-NAU3481	3.43	12.2
	592	<i>qVV-D11-1BC₁S₂592</i>	D11	NAU1640-BNL3279	3.40	22.3

^aMarkers flanking the QTL likelihood peak.

was associated with *qVV-A8-IBCIS2BP2* and *qVL-A8-IBCIS2VD8* on Chromosome A8. Markers linked to the Chromosome 11 QTL (Bolek et al 2005) mapped to the homoeologous chromosome (D11) and was linked to *qVV-D11-IBCIS2VD8* (Yang et al. 2008).

5 Reverse and Comparative Genetics

5.1 Expressed Sequence Tags/Genome-Wide Expression

Recently, a community-wide effort produced 185,198 Expressed Sequence Tags (ESTs) from 30 cDNA libraries, sampling a variety of tissues and developmental stages, including several subjected to abiotic stresses such as chilling temperatures and water-deficit (www.agcol.arizona.edu/pave/cotton/) (Udall et al. 2006). Subsequently, a subset of these sequences was used to generate a publicly available oligonucleotide microarray for gene expression studies. Even though this array does not provide complete transcriptome coverage, particularly for pathogen response genes, it has provided a foundation for more robust functional analysis of a multitude of developmental responses in cotton. Additionally, the ability to integrate expressed sequence data with structural genomic data may allow the identification of functional regions on chromosomes that contribute to variability in quantitative traits.

An NSF project IOB-0090920 (PIs Essenberg and Pierce) has created important genetic resources for the functional analysis of genes during the hypersensitive response of *Xcm* infection. This includes a series of isogenic lines (Acala 44) containing the *B*₂, *B*₄, *B*₅, *B*₆, *b*₇, and *B*_m loci (Essenberg, Bayles, Samad, Hall, Brinkerhoff and Verhalen 2002) and approximately 2,000 EST developed from a subtracted, normalized library enriched for sequences induced during the hypersensitive response of *Xcm*. They have created a DNA microarray which is currently being used to investigate and compare expression profiles among the above series of isogenic lines.

5.2 Resistance Gene Analogues (RGA)

Resistance genes cloned from several plant species to date have demonstrated a few recurring themes. Many can be classified into groups that encode a nucleotide binding site (NBS) and leucine rich repeats (LRR) (Collins, Webb, Seah, Ellis, Hulbert and Pryor 1998; Ellis, 1998; Martin 1999; Ellis, Dodds and Pryor 2000). This NBS-LRR class of resistance genes is abundant in many plant species, including flax (*L6*), *Arabidopsis* (*RPS2*), and tobacco (*N*) (Bent, Kundel, Dahlbeck, Brown, Schmidt, Giraudat, Leung and Staskawicz 1994; Whitham, Dinesh-Kumar, Choi, Hehl, Corr and Baker 1994; Lawrence, Finnegan, Ayliffe and Ellis 1995). Two subclasses of NBS-LRR exist, one

that also contains an amino-terminal Toll/interlucan-1-receptor homology region (TIR) and another without the TIR region. The *Xa21* resistance gene in rice (*Xanthomonas* resistance) or the *Cf* genes in tomato (*Cladosporium fulvum* resistance) are classified as extracellular LRR (Ohmori, Murata and Motoyoshi 1998). The first positionally cloned resistance gene (*Pto*) encodes a serine/threonine protein kinase (PK) and requires the *Prf* (Pseudomonas resistance and fenthion sensitivity) gene for activity. These classes of genes have been implicated in gene-for-gene resistance to fungal, viral, bacterial and nematode pathogens (Collins et al. 1998; Ellis, 1998; Martin, 1999; Ellis et al. 2000). In many species, oligonucleotide primers designed from these conserved coding regions have been used to amplify related sequences in other species.

No less than 150 resistance gene analogues (RGAs) have been isolated and characterized in cotton (Tan, Callahan, Zhang, Karaca, Saha, Jenkins, Creech and Ma 2003; He, Du, Covalada, Xu, Robinson, Yu, Kohel, and Zhang 2004; Hinchliffe, Lu, Potenza, Segupta-Gopalan, Cantrell, and Zhang 2005). The phylogenetic relationship of RGA can be put into 11 groups and 56 sub-groups based on genetic distances. A molecular marker system that combines attributes of amplified fragment length polymorphism (AFLP) with a degenerate RGA designed primer offers a marker platform to analysis many putative RGA loci simultaneous (Zhang, Yuan, Niu, Hinchliffe, Lu, Yu, Percy, Ulloa, and Cantrell 2007).

6 Molecular Breeding: Bridging Elite and Underutilized Germplasm Pools

Breeding of modern cotton cultivars is limited to a very narrow set of R-genes (alleles) that were present in the small number of domesticated genotypes or subsequently introduced from other germplasm sources. While it is widely recognized that exotic cotton germplasm contains potentially valuable sources of disease resistance, the difficulties associated with separation of these valuable R-genes from nearby undesirable genes have deterred most breeders from pursuing such efforts. In only few cases have breeders successfully used feral *Gossypium* germplasm to reduce vulnerability to key pathogens.

While tetraploid cotton is reproductively incompatible with diploid cotton (Beasley 1942), the transfer and deployment of diploid R-genes was successfully accomplished in cotton. Knight successfully transferred bacterial blight resistance genes from diploid *G. arboreum* into tetraploid cotton by creating a synthetic tetraploid followed by successive back crossing to the tetraploid parent (Knight 1948, 1953; Knight and Hutchinson 1950). These R-genes have historically been a very important source of resistance to the most virulent *Xcm* races, including Race 18 (Bird 1982; Wright et al. 1998). Many modern breeding programs around the world have used these R-genes to minimize potential bacterial blight epidemics.

Cotton breeding is rapidly shifting from traditional phenotypic selection to genetic technologies that enable the direct selection and examination of genes or alleles. The implications of this shift for breeding and using novel germ-plasm sources have recently been realized. The development of resistance to reniform nematodes (*Rotylenchulus reniformis*) from F-genome diploid *G. longicalyx* (Robinson et al. 2007) used a strategy that merged new and traditional technologies to accelerate the discovery and transfer of resistance into elite cotton. The contributions that these scientists have made should be applauded.

Corporate molecular breeding activities currently focused on gene discovery and predictive breeding of resistant cultivars have immense potential to capture market share. The proprietary and breeding value of novel resistant genotypes has led to an increased investment by private industry.

7 Into the Future

The assessment of a plant's response to pathogen infection is a key step in dissecting host-pest interactions. The use of inoculated and natural infections to elicit a symptomatic response can provide mixed results. It is clearly evident from the body of research that a single-plant phenotypic evaluation is less than 100% accurate and that some susceptible individuals will lack disease symptoms. When designing an experiment, the accuracy, repeatability and through-put of the phenotypic assay are important considerations. The use of appropriate controls (positive and negative) and a well defined contingency plan to address potential pitfalls are also important factors to consider. Clearly the dynamics of a segregating population will compound problems associated with a phenotypic assay. Activities should be under-taken to maximize meaningful results by inoculating plants in controlled conditions, progeny testing individuals, including appropriate controls, and using plant materials applicable to replicated testing.

Contributions from cotton breeders, geneticists, and pathologists have provided a better understanding of the plant-pathogen interactions leading to disease symptoms. Key advancements from this research include the identification of the pathogenic organisms, pathotypes or races, sources of resistant germplasm and host differentials. In the past two decades much advancement at the cellular and molecular level have led to better comprehension of the disease cycle, plant defense responses, and genes involved in cotton disease and host resistance. Continued advances in molecular genetics offer many tools useful for the examination of genetic and biochemical mechanisms controlling plant-pathogen interaction. Cotton is a vital global commodity with economic importance to many countries; continued research focused on disease will remain a high-priority.

References

- Ahuja, S.L., Monga, D. and Dhayal, L.S. (2007) Genetics of Resistance to Cotton Leaf Curl Disease in *Gossypium hirsutum* L. under Field Conditions. *J. of Heredity* 98, 79–83.
- Ali, M. (1997) Breeding of cotton varieties for resistance to cotton leaf curl virus. *Pakistan Journal of Phytopathology* 9, 1–7.
- Aslam, M., Jiang, C., Wright, R.J. and Paterson, A.H. (1999) Identification of Molecular Markers Linked to Leaf Curl Virus Disease Resistance in Cotton. *Pakistan J. Biol. Sci.* 2, 124–126.
- Assigbetse, K.B., Fernandez, D., Dubois, M.P. and Geiger, J.P. (1994) Differentiation of *Fusarium oxysporum* f. sp. *vasinfectum* races on cotton by random amplified polymorphic DNA (RAPD) analysis. *Phytopathology* 84, 622–626.
- Barrow, J.R. (1970a) Heterozygosity in inheritance of Verticillium wilt tolerance in cotton. *Phytopathology* 60:301–303.
- Barrow, J.R. (1970b) Critical requirements for genetic expression of Verticillium wilt tolerance in Acala cotton. *Phytopathology* 60:559–560.
- Beasley, J.O. (1942) Meiotic chromosome behavior in species, species hybrids, haploids and induced polyploids of *Gossypium*. *Genetics* 27, 25–54.
- Bell, A.A. (1994) Mechanisms of disease resistance in *Gossypium* species and variation in *Verticillium dahliae*. In: Constable GA, Forrester NW (eds), *Challenging the Future: Proceedings of the World Cotton Research Conference-1*, February 14–17, Brisbane, Australia, pp. 225–235.
- Bird, L.S. (1976) Registration of Tamcot SP21, Tamcot SP23 and Tamcot SP37 cottons. *Crop Science* 16, 884.
- Bird, L.S. (1982) The MAR (multi-adversity resistance) system for genetic improvement of cotton. *Plant Dis.* 66, 172–176.
- Blasingame, D. (1995/2006 separate volumes) Cotton Disease Loss Estimate. Committee Report. Proc. Beltwide Cotton Conf., Cotton Disease Council, National Cotton Council, Memphis, Tenn.
- Bolek, Y., El-Zik, K.M., Pepper, A.E., Bell, A.A., Magill, C.W., Thaxton, P.M. and Reddy, O.U.K. (2005) Mapping of Verticillium wilt resistance genes in cotton. *Plant Sci.* 168, 1581–1590.
- Bent, A.F., Kundel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J. and Staskawicz, B.J. (1994) *RPS2* of *Arabidopsis thaliana*: a Leucine-rich repeat class of plant disease resistance gene. *Science* 265:1856–1860.
- Bridge, P.D., Ismail, M.A. and Rutherford, M.A. (1993) An assessment of aesculin hydrolysis, vegetative compatibility and DNA polymorphism as criteria for characterizing pathogenic races within *Fusarium oxysporum* f. sp. *vasinfectum*. *Plant Pathol.* 42, 264–269.
- Collins, N.C., Webb, C.A., Seah, S., Ellis, J.G., Hulbert, S.H. and Pryor A. (1998) The isolation and mapping of disease resistance gene analogs in Maize. *MPMI* 11:968–978.
- Davis, R.D., Moore, N.Y. and Kochman, J.K. (1996) Characterisation of a population of *Fusarium oxysporum* f. sp. *vasinfectum* causing wilt of cotton in Australia. *Australian Journal of Agriculture Research* 47:1143–1156.
- Davis, R.M., Colyer, P.D., Rothrock, C.S. and Kochman, J.K. (2006) Fusarium Wilt of Cotton: Population Diversity and Implications for Management. *Plant Disease* 90, 692–703.
- De Feyter, R. and Gabriel, D.W. (1991) At least six avirulence genes are clustered on a 90-kilobase plasmid in *Xanthomonas campestris* pv. *malvacearum*. *Mol. Plant Microbe Interact.* 4:423–432.
- De Feyter, R., McFadden, H. and Dennis, L (1998) Five avirulence genes from *Xanthomonas campestris* pv. *malvacearum* cause genotype-specific cell death when expressed transiently in cotton. *Mol. Plant Microbe Interact.* 11:698–701

- Delannoy, E., Lyon, B.R., Marmey, P., Jalloul, A. Daniel, J.F., Montillet, J.L. Essenberg, M. and Nicole, M. (2005) Resistance of Cotton Towards *Xanthomonas campestris* pv. *Mala-vacearum* Annu. Rev. Phytopathol. 43, 63–82.
- Devey, M.E. and Roose, M.L. (1987) Genetic analysis of *Verticillium* wilt tolerance in cotton using pedigree data from three crosses. Theor. Appl. Genet. 74, 162–167.
- Du, W.S., Du, X.M. and Ma Z.Y., (2004) Studies on SSR markers of resistance gene of *Verticillium* wilt in cotton. J. Northwest Sci. Tech. Univ. Agric.: Nat. Sci. Ed. 32, 20–24.
- Ellis, J. (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr. Opin. Plant Biol. 1:288–293.
- Ellis, J., Dodds, P., and Pryor, T. (2000) Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. 3:278–284.
- El-Zik, K.M. and Bird, L.S. (1970) Effectiveness of specific genes and gene combinations in conferring resistance to races of *Xanthomonas malvacearum* in Upland cotton. Phytopathology 60, 441–447.
- Essenberg, M, Bayles, M.B., Samad, R.A., Hall, J.A., Brinkerhoff, L.A. and Verhalen, L.M. (2002) Four new-isogenic lines of cotton with different genes for bacterial blight resistance. *Phytopathology* 92:1323–1328.
- Fernandez, D., Assigbetse, K., Dubois, M.P. and Geiger, J.P. (1994) Molecular characterization of races and vegetative compatibility groups in *Fusarium oxysporum* f. sp. *vasinfectum*. Appl. Environ. Microbiol. 60, 4039–4046.
- Follin, J.C., Girardot, B., Mangano, V. and Benitez, R. (1988) New results on inheritance of immunity to bacterial blight (*Xanthomonas campestris* pv. *malvacearum* (Smith) Dye, race 18 and 20) in the cotton plant (*Gossypium hirsutum* L.). Coton et Fibres Tropicales 43, 167–175.
- Gabriel, D.W. (1999) The *Xanthomonas avr/pth* gene family. In *Plant-Microbe Interactions*, ed. G Stacey, NT Keen, 4:39–55. St Paul, MN: APS Press
- Gao, Y.Q., Nie, Y.C. and Zhang, X.L. (2003) QTL mapping of genes resistant to *Verticillium* wilt in cotton. Cotton Sci. 15: 73–78.
- Hameed, S., Khalid, S., Ehsan-ul-Haq, Hashrni, A.A. (1993) Cotton leaf curl disease in Pakistan caused by a whitefly-transmitted geminivirus. Plant Dis 78: 529
- He L., Du, C., Covalada, L., Xu, Z., Robinson, A.F., Yu, J.Z., Kohel, R.J., and Zhang H-B. (2004) Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). Mol Plant–Microbe Interact 17:1234–1241
- Hillocks, R.J. (1983) Infection of cotton seed by *Fusarium oxysporum* f.sp. *vasinfectum* in cotton varieties resistant or susceptible to Fusarium wilt. Trop. Agric. (Trinidad) 60, 141–143.
- Hillocks, R.J. (1992) *Cotton Disease*. CAB International, Wallingford, UK.
- Hillocks, R.J. and Kibani, T.H.M (2002) Factors affecting the distribution, incidence and spread of Fusarium wilt of cotton in Tanzania. Expl. Agric. 38, 13–27.
- Hinchliffe, D. J., Lu, Y. Z., Potenza, C., Segupta-Gopalan, C., Cantrell, R. G. and Zhang, J. F. (2005) Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. Theor. Appl. Genet. 110:1074–1085
- Innes, N.L. (1983) Bacterial blight of cotton. *Biol. Rev.* 58, 157–176.
- Kim, Y., Hutmacher, R.B. and Davis, R.M. (2005) Characterization of California Isolates of *Fusarium oxysporum* f. sp. *vasinfectum*. Plant Dis 89, 366–372.
- Kirkpatrick, T.L. and Rothrock, C.S. (2001) *Compendium of Cotton Disease*. The American Phytopathological Society, St. Paul, Minnesota.
- Knight, R.L. (1948) The role of major genes in the evolution of economic characters. J. Genet. 48: 370–387.
- Knight, R.L. (1953) The genetics of blackarm resistance. IX. The gene $B_6 M$ from *Gossypium arboreum*. J. Genet. 51, 270–275.
- Knight, R.L. (1963) The genetics of blackarm resistance. XII. Transference of resistance from *Gossypium herbaceum* to *Gossypium barbadense*. Genetics 50, 36–58.

- Knight, R.L. and Hutchinson, J.B. (1950) The evolution of blackarm resistance in cotton. *J. Genet.* 50: 36–58.
- Lawrence, G.J., Finnegan, E.J., Ayliffe, M.A., and Ellis, J.G. (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene RPS2 and the tobacco viral resistance gene *N*. *Plant Cell* 7:1195–1206.
- Leach, J.E. and White, F.F. (1997) Avirulence genes. In *Plant-Microbe Interactions*, ed. G. Stacey, NT Keen, 2:61–98. New York: Chapman & Hall
- Luckett, D.J., Thomson N.J. and Reid P.E. (1986) Quality requirements and cotton breeding in Australia. In *Plant Breeding Symposium*. Lincoln, New Zealand.
- Martin, G.B. (1999) Functional analysis of plant disease resistance genes and their downstream effectors. *Curr. Opin. Plant Biol.* 2:273–279.
- Mansoor, S., Briddon, R.W., Bull, S.E., Bedford, I.D., Bashir, A., Hussain, M., Saeed, M., Zafar, Y., Malik, K.A., Fauquet, C., Markham, P.G. (2003) Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA β . *Archives of Virology* 148, 1969–1986.
- McFadden, H., Beasley, D. and Brubaker, C.L. (2004) Assessment of *Gossypium sturtianum* and *G. australe* as potential sources of Fusarium wilt resistance to cotton. *Euphytica* 138:61–72.
- Mert, M., Kurt, S., Gencer, O., Akiscan, Y., Boyaci, K. and Tok, F.M. (2005) Inheritance of resistance to *Verticillium* wilt (*Verticillium dahlia*) in cotton (*Gossypium hirsutum* L.). *Plant Breeding* 124, 102–104.
- Nelson, R.R. (1978) Genetics of horizontal resistance to plant diseases. *Annu. Rev. Phytopathol.* 16, 359–378.
- Niu, C., Hinchliffe, D.J., Cantrell, R.G., Wang, C., Roberts, P.A. and Zhang, J.F. (2007) Identification of molecular markers associated with root-knot nematode resistance in Upland cotton. *Crop Sci.* 47, 951–960.
- Niu, C., Lister, H.E., Nguyen, B., Wheeler, T.A. and Wright, R.J. (2008) Quantitative trait loci controlling resistance to *Thielaviopsis basicola* in diploid cotton. *Theor. Appl. Genet.* 117, 1313–1323.
- Ohmori T., Murata, M. and Motoyoshi, F. (1998) Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. *Theor. Appl. Genet.* 96, 331–338.
- Pan, J.J., Zhang, T.Z., Kuai, B.K., Guo, X.P. and Wang, M. (1994) Studies on the inheritance of resistance *Verticillium dahlia* in cotton. *J. Nanjing Agric. Univ.* 17, 8–18.
- Patil, M.A, Pierce, M.L., Phillips, A.L., Venters, B.J. and Essenberg, M. (2005) Identification of genes up-regulated in bacterial-blight-resistant upland cotton in response to inoculation with *Xanthomonas campestris* pv. *malvacearum*. *Physiological and Molecular Plant Pathology* 67: 319–335.
- Rahman, M., Hussain, D., Malik, T.A. and Zafar, Y. (2005) Genetics of resistance to cotton leaf curl disease in *Gossypium hirsutum*. *Plant Pathology* 54, 764–772.
- Roberts, C.L. and Staten, G. (1972) Heritability of *Verticillium* wilt tolerance in crosses of American upland cotton. *Crop Sci.* 12, 63–66.
- Robinson, A.F., Bell, A.A., Dighe, N.D., Menz, M.A., Nichols, R.L. and Stelly D.M. (2007) Introgression of resistance to nematode *Rotylenchulus reniformis* into Upland Cotton (*Gossypium hirsutum*) from *Gossypium longicalyx*. *Crop Sci.* 47, 1865–1877.
- Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P-W, Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C-H, Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schultze, S.R., Trolinder, N.L., Wendel, J.F., Wilkins, T.A., Williams-Coplin, D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. and Paterson, A.H. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*: 166, 389–417
- Rong J., Bowers, J.E., Schulze, S.R., Waghmare, V.N., Rogers, C.J., Pierce, G.J., Zhang, H., Estill, J.C., Paterson, A.H. (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Res.* 15:1198–1210

- Rong, J., Feltus, F.A., Waghmare, V.N., Pierce, G.J., Chee, P.W., Draye, X., Saranga, Y., Wright, R.J., Wilkins, T.A., May, O.L., Smith, C.W., Gannaway, J.R., Wendel, J.F. and Paterson, A.H. (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics* 176:2577–2588
- Rungis, D., Llewellyn, D., Dennis, E.S. and Lyon, B.R. (2002) Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. *Aust. J. Agric. Res.* 53, 551–560
- Saunders, J.H. and Innes, N.L. (1963) The genetics of bacterial blight resistance in cotton: Further evidence on the gene *B_{6m}*. *Genet. Res. Camb.* 4, 382.
- Shen, X., Becelaere, G. V., Kumar, P., Davis, R. F., May, O. L. and Chee, P. (2006) QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. *Theor. Appl. Genet.* 113, 1539–1549.
- Skovgaard, K., Nirenberg, H.I., O'Donnell, K. and Rosendahl, S. (2001) Evolution of *Fusarium oxysporum* f. sp. *vasinfectum* races inferred from multigene genealogies. *Phytopathology* 91:1231–1237.
- Taliercio, E., Allen, R.A., Essenberg, M., Klueva, N., Nguyen, H., Patil, M.A., Payton, P., Millena, A.C.M., Phillips, A.L., Pierce, M.L., Scheffler, B., Turley, R., Wang, J., Zhang, D. and Scheffler, J. (2006) Analysis of ESTs from multiple *Gossypium hirsutum* tissues and identification of SSRs. *Genome* 49, 306–319.
- Tan H., Callahan, F.E., Zhang, X-D, Karaca, M., Saha, S., Jenkins, J.N., Creech, R.G. and Ma, D-P. (2003) Identification of resistance gene analogs in cotton (*Gossypium hirsutum* L.). *Euphytica* 134:1–7
- Udall, J.A., Swanson, J.M., Haller, K., Rapp, R.A., Sparks, M.E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A.B., Sickler, B.A., Wilkins, T.A., Guo, J.Y., Chen, X.Y., Scheffler, J., Taliercio, E., Turley, R., McFadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S.R., Pierce, M.L., Essenberg, M., Kim, H., Llewellyn, D.J., Dennis, E.S., Kudrna, D., Wing, R., Paterson, A.H., Soderlund, C. and Wendel, J.F. (2006) A global assembly of cotton ESTs. *Genome Res.* 16: 441–450.
- Ulloa, M., Hutmacher, R. B., Davis, R. M., Wright, S. D., Percy, R. and Marsh, B. (2006) Breeding for Fusarium Wilt Race 4 Resistance in Cotton under Field and Greenhouse Conditions. *J. Cotton Sci* 10, 114–127.
- Verhalen, L.M., Brinkerhoff, L.A., Fun, K.C. and Morrison, W.C. (1971) A quantitative genetic study of *Verticillium* wilt resistance among selected lines of upland cotton. *Crop Sci.* 11, 407–412.
- Wang, H.M, Zhang, X.L., He, D.H., Lin, Z.X., Nie, Y.C., Li, Y.H. and Chen W. (2005) Detection of DNA markers associated with resistance to *Verticillium dahlia* in cotton. *Acta Phytopathol. Sin.* 34, 333–339.
- Wheeler, T. A., Gannaway, J. R. and Keating, K. (1999) Identification of resistance to *Thielaviopsis basicola* in diploid cotton. *Plant Dis.* 83, 831–833.
- Wheeler, T. A. and Gannaway, J. R. (2007) Identification of germplasm resistant to *Thielaviopsis basicola* in the USDA cotton germplasm collection. In *World Cotton Conference-4*, 10–14 Sept, 2007, Lubbock, TX. (In press)
- Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C. and Baker, B. (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to the Toll and the interleukin-1 receptor. *Cell* 78:1101–1115.
- Wright, R.J., Thaxton, P.M., El-Zik, K.M. and Paterson, A.H. (1998) D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics* 149, 1987–1996
- Yang, C., Guo, W., Li, G., Gao, F., Lin, S. and Zhang, T. (2008) QTL mapping of *Verticillium* wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Science* 174, 290–298.

- Yang, Y. and Gabriel, D.W. (1995) *Xanthomonas* avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Mol. Plant Microbe Interact.* 8, 627–631.
- Yang, Y., Yuan, Q. and Gabriel, D.W. (1996) Watersoaking function(s) of *XcmH1005* are redundantly encoded by members of the *Xanthomonas avr/pth* gene family. *Mol. Plant Microbe Interact.* 9:105–113
- Zhang, J., Yuan, Y., Niu, C., Hinchliffe, D.J., Lu, Y., Yu, S., Percy, R.G., Ulloa, M. and Cantrell, R.G. (2007) AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. *Crop Science* 47:180–187
- Zhen, R., Wang, X.F., Ma, Z.Y., Zhang, G.Y. and Wang, X. (2006) A SSR marker linked with the gene of *Verticillium* wilt resistance in *Gossypium barbadense*. *Cotton Sci.* 18, 269–272.

Bridging Classical and Molecular Genetics of Abiotic Stress Resistance in Cotton

Yehoshua Saranga, Andrew H. Paterson, and Avishag Levi

Abstract The effect on abiotic stress on productivity of crop plants and particularly cotton is discussed (section 1) with an emphasis on water stress, the major abiotic stress factor (section 2). Plant responses to abiotic stress are briefly discussed (section 3), followed by the origin of cotton and its implications for the available genetic resources for abiotic stress resistance (section 4). The next sections describe attempts to dissect and improve cotton resistance to abiotic stress using classical genetics (section 5), genomic (section 6) and transgenic (section 7) approaches. It is concluded (section 8) that there is an urgent need to improve abiotic stress resistance of cotton, a task that has become more feasible with the currently available knowledge and genomic tools.

1 Effect of Abiotic Stress on Crops Productivity

Crop plants are often grown under unfavorable environmental conditions that prevent the full expression of their genetic potential for productivity. Boyer (1982) reported that the average production of major U.S. crops (corn, wheat, soybean, sorghum, oat, barley, potato and sugar beat) is only 21.6% of the highest yields attained. Disease and insect losses account for 4.1% and 2.6% yield reductions, respectively, with the remainder of the yield reduction attributed to unfavorable physicochemical (abiotic) environments induced by weed competition (2.6%) and inappropriate soils and climates (69.1%). Similar figures were presented recently for corn, showing that yield losses to all stresses sum to 71% in developed temperate areas and 80% in less developed tropical areas, with most of the losses attributed to abiotic factors (59% and 64%, in temperate and tropical areas, respectively) (Edmeades, Cooper, Lafitte, Zinselmeier, Ribaut, Habben, Loffler and Banziger 2001). Certainly, some of these

Y. Saranga (✉)

The RH Smith Institute of Plant Science and Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot 76100, Israel
e-mail: saranga@agri.huji.ac.il

losses are caused by inherently unfavorable environments and some by sub-optimal management practices by farmers, often due to economic reasons. Nevertheless, there is no doubt that a large fraction of potential crop productivity is lost to abiotic stress factors.

Cotton (*Gossypium hirsutum* L. and *G. barbadense* L., abbreviated hereafter as *GH* and *GB*, respectively) is the world's leading fiber crop and also an important oilseed. In Israel, where the cotton crop benefits from a long summer, favorable temperatures and high inputs (drip irrigated, fertigation, pest control etc.), the average country-wide yield of Upland cotton (*GH*) in 2005 was 2020 Kg lint ha⁻¹ with specific commercial fields producing as much as 2400 Kg lint ha⁻¹ (Israel Cotton Board, personal communication); similar record yields are obtained also in Australia under comparable conditions. At the same time world average cotton yield is estimated at 650 Kg lint ha⁻¹ (FAOSTat, <http://faostat.fao.org/default.aspx>), reflecting a 73% loss to various stresses - comparable to those values published for other crops (Boyer 1982; Edmeades et al. 2001).

2 Water Deficiency - a Major Abiotic Stress Factor

Among the various abiotic stress conditions, (water – deficit or access, temperature – low or high, ions - deficit or access, light - deficit or access) water deficiency is the most devastating factor (Boyer 1982; Araus et al., 2002). About 44.9% of US soils are subjected to water deficiency due to either drought (25.3%) or shallowness (19.6%), with an additional 43% subjected to other environmental limitations, primarily cold (16.5%) and water logging (15.7%) (Boyer 1982). Only 12.1% of US soils are considered free from physicochemical problems. In agreement with this, of the total indemnification made to U.S. farmers for crop losses due to drought, excessive water and cold account for 40.8%, 16.4% and 13.8%, respectively (Boyer 1982).

Cotton is among the major herbaceous warm-season crops, grown during the summer in arid and semiarid regions where water availability is often limited, but also in temperate regions up to 47° N latitude (Singh, Prasad, Sunita, Giri and Reddy 2007). As such, water deficit is the major abiotic stress limiting cotton productivity (Pettigrew 2004), but it is also considerably affected by salinity (Ashraf 2002), and extreme temperatures (Singh et al. 2007). The high value of cotton justifies irrigation, and accordingly it is a major consumer of agricultural water. Regardless of whether it is irrigated or not, cotton is often exposed to drought, which adversely affects both yield and lint quality.

Currently, about 80% of the world's useable water resources are consumed by irrigated agriculture (Condon, Richards, Rebetzke and Farquhar 2004). One-third of the world's arable land suffers from chronically inadequate water availability for agriculture, and in virtually all agricultural regions, crop yields are periodically reduced by drought (Kramer 1980; Bruce, Edmeades and Barker 2002). Within a few decades, the constantly expanding world population will require more water for domestic, municipal, industrial, and environmental

needs (Hamdy, Ragab and Scarascia-Mugnozza 2003), as well as larger amounts of food and fiber. To meet these projected demands, more efficient production from less irrigation water will be required (Condon et al. 2004). This trend is expected to accentuate due to global climatic change and increased aridity (Vörösmarty, Green, Salisbury and Lammers 2000). Therefore, developing stress resistant crop plants is vital to meet increased demand for food and fiber (Plucknett, Smith, Williams and Anishetty 1987; Parry, Flexas and Medrano 2005). Improving crop stress resistance requires comprehensive exploration of available genetic resources and an integrative understanding of their adaptive mechanisms at the physiological, genetic, and genomic levels. Understanding plant responses to water deficit is important not only due to being the major stress affecting crop productivity, but also because other stresses associate with (heat) or induce (cold and salinity) water deficiency in plant tissues (Bohnert, Nelson and Jensen 1995). Therefore, in view of its abundance, impact and expected expansion, major emphasis will be placed in this chapter on water deficiency.

3 Plant Responses to Abiotic Stress

Resistance of plants to stress conditions may arise from escape, avoidance, or tolerance strategies (Levitt 1972; Turner 1986). **Escape** relies on successful reproduction before the onset of severe stress (i.e. developmental plasticity) (Mooney, Percy and Ehleringer 1987; Maroco, Petreira and Chaves 2000). **Avoidance** involves the prevention or decreased penetration of the stress into the plant tissues, such as minimizing water loss and maximizing water uptake (Chaves, Maroco and Pereira 2003) or exclusion of salt ions (Munns, James and Lauchli 2006). **Tolerance** relies on resisting the stress in spite of permitting it to enter the plant tissues. This strategy involves co-ordination of physiological and biochemical alterations at the cellular and molecular levels, such as osmotic adjustment (Morgan 1984) and the absorption of salt ions and their sequestration in the vacuole (Mimura, Kura-Hotta, Tsujimura, Ohnishi, Miura, Okazaki, Mimura, Maeshima and Washitani-Nemoto 2003). In most cases, plants subjected to stress conditions combine a range of response types, showing a number of physiological responses at the molecular, cellular, and whole-plant levels (Bray 1993; Bartels, Furini, Ingram and Salamini 1996; Chaves et al. 2003).

Two stress-related physiological mechanisms warrant brief mention in this framework. (i) **Water use efficiency** (WUE) is defined in agronomic terms as the ratio between dry matter (DM) production and water use (Condon and Hall 1997). In physiological terms, however, WUE is defined as the ratio between the rate of carbon fixed and the rate of water transpired. Carbon isotope ratio ($^{13}\text{C}/^{12}\text{C}$, expressed with differential notation as $\delta^{13}\text{C}$) is commonly used as an indirect indicator of WUE (Farquhar, O'Leary and Berry 1982; Araus, Bort, Steduto, Villegas and Royo 2003). (ii) **Osmotic adjustment** is a net increase in

intercellular solutes in response to water stress (Morgan 1984). Osmotic adjustment is considered one of the crucial processes in plant adaptation to drought and salinity, because it allows turgor maintenance at lower water potential, sustains tissue metabolic activity and enables regrowth upon rewatering. Plant productivity under arid conditions has been associated with osmotic adjustment in a number of crop species (reviewed by Blum 2005), including cotton (Saranga, Menz, Jiang, Wright, Yakir and Paterson 2001; Saranga, Jiang, Wright, Yakir, and Paterson 2004).

In the field, crops plants are routinely subjected to a combination of different abiotic stresses. Plant responses to one stress can be synergistically or antagonistically modified by the superimposition of other stresses. Water loss from a plant (transpiration) is an unavoidable consequence of photosynthesis (Cowan 1986), whereby the energy of solar radiation is used for carbon fixation. While increased transpiration without a corresponding increase in photosynthesis reduces WUE it is also a benefit in dissipating excess heat (Cornish, Radin, Turcotte, Lu and Zeiger 1991; Radin, Lu, Percy and Zeiger 1994). Water stress and heat stress almost invariably co-occur under arid-region field conditions. The resulting need for a balance between tolerance of heat and drought complicates strategies for manipulating plant water use to improve productivity under arid conditions. Another example of interaction between two abiotic stresses occurs under salinity. The immediate effect of salinity, reduced soil osmotic potential, is similar to water stress and only its later effect, ion toxicity, is specific to salt (Munns 2005). Moreover, primary abiotic stresses, such as drought, salinity, cold, heat and chemical pollution, are often interconnected with secondary oxidative stress by inducing the production of reactive oxygen species (ROS) in plants (Vinocur and Altman 2005).

4 The Origin of Cotton and its Implications for Abiotic Stress Resistance

As discussed in further detail in other chapters of this volume, the *Gossypium* genus includes about 50 diploid ($2n = 26$) and 5 tetraploid ($2n = 4x = 52$) species (Fryxell 1992). All tetraploid cottons are comprised of 'A' and 'D' subgenomes that diverged from a common ancestor about 4–11 million years ago then were rejoined in a common nucleus about 1–2 million years ago (Wendel 1989). Virtually all genes in tetraploid cotton are represented by two or more copies, with one in each of the two subgenomes (Rong, Abbey, Bowers, Brubaker, Chang, Chee, Delmonte, Ding, Garza, Marler, Park, Pierce, Rainey, Rastogi, Schulze, Trolinder, Wendel, Wilkins, Williams-Coplin, Wing, Wright, Zhao, Zhu and Paterson 2004) and in similar (albeit not identical) chromosomal orders in the two tetraploid subgenomes and their diploid ancestors (Brubaker, Paterson and Wendel 1999). The tetraploid species *GH* and *GB* include cultivated forms that dominate cotton commerce. The cultivated cotton

species can be crossed to one another, or to any of the three wild tetraploid species, *G. tomentosum*, *G. darwinii* and *G. mustelinum* to yield hybrids that show normal meiosis and produce many fertile progenies, albeit with varying degrees of sterility and incompatibility of some interspecific allele combinations leading to non-random patterns of introgression (Jiang, Chee, Draye, Morrell, Smith and Paterson 2000; Waghmare, Rong, Rogers, Pierce, Wendel and Paterson 2005).

Cotton originates from wild perennial plants adapted to semi-arid, semi-tropical environments which experienced periodic drought and temperature extremes (Kohel 1974). Therefore, adaptations to heat, drought and salt stresses are expected to exist within the *Gossypium* genus. Indeed, exotic tetraploid cottons are well-adapted to heat and drought stress. *G. tomentosum* is endemic to Hawaii, where it is found in arid, rocky, or clay coastal plains on the leeward side of all main islands except the Big Island (http://kalama.doe.hawaii.edu/hern95/pt009/Ann/mcc_nativeplants.html), and the progenies of an interspecific cross between *G. tomentosum* and *GH* showed a high degree of tolerance to drought (Gotemare and Singh 2004). *G. mustelinum* is endemic to extremely dry areas (Pickergill, Barrett and de Andrade-Lima 1975) in a few states of the semi-arid region of Northeastern Brazil (Silva, Camara, Medeiros, Oliviera, Agra, Harley and Giulietti 2006). *G. darwinii* is endemic to the Galapagos islands, exhibiting the “... capacity to survive naturally in the exacting environment of those arid islands” (Fryxell 1979). An exotic strain of *GH*, var. *marie-galante*, was cultivated in Brazil in the beginning of the 20th century due to its superior drought tolerance (Boulanger and Pinheiro 1971). Quisenberry, Jordan, Roark and Fryrear 1981) showed variation in dry matter accumulation, heat tolerance, root growth, and WUE among exotic *GH* lines. Wild *GB* plants grow on the coasts of Peru and Ecuador and perhaps the Galapagos Islands, subjected to salinity stress (Lee 1984). In summary, the large number of wild *Gossypium* species presents an impressive range of variation in many characters, as do the numerous variants within the cultivated species, all of which is potentially available for exploitation in cotton improvement programs (Fryxell 1984). Therefore, a major objective in modern cotton breeding is to restore from wild ancestors of crop plants alleles that were “left behind” during domestication and may be useful for the improvement of productivity under stress conditions (Tanksley and McCouch 1997; Gur and Zamir 2004).

5 Classical Genetics of Cotton Responses to Abiotic Stress

The genetic bottleneck imposed on crop plants by domestication and modern plant-breeding has severely eroded the allelic variations of genes originally found in the wild (e.g. Tanksley et al. 1997). Modern cotton cultivars are products of intensive selection to facilitate mechanical harvesting and processing as well as to produce large amounts of specific types of fibers. This selection

has narrowed the genetic variability for abiotic stress resistance (Rosenow, Quisenberry, Wendt and Clark 1983). However, considerable variation persists within and between the two dominant cultivated cotton species, *GH* and *GB*, for physiological traits that confer resistance to drought (Yakir, De Niro and Ephrath 1990; Saranga, Flash and Yakir 1998), salinity (Ashraf and Ahmed 2000) and high temperatures (Singh et al. 2007).

Classical cotton breeding made a major contribution to the advancement of the crop. A study of cotton cultivars released between 1910 and 1980 showed a continues yield gain by an average rate of 10.4 kg/ha year⁻¹ (Meredith and Bridge 1984). Achievements were made also in breeding for biotic stress (pest and disease) resistance and the development of early maturing cultivars. We are not aware, however, of a major breeding effort targeted towards improving cotton resistance to abiotic stresses. Most attempts to improve cotton adaptation to abiotic stress conditions seemed to rely mainly on screening for stress resistance of existing cultivars.

Advances in the yield of Pima cotton (*GB*) released during the second half of the 20th century have been associated with heat avoidance via increased stomatal conductance (Cornish et al. 1991; Radin et al. 1994). Lint yield was positively correlated with stomatal conductance (Radin et al. 1994), which in turn correlated negatively with $\delta^{13}\text{C}$ (Lu, Chen, Percy, Sharifi, Rundel and Zeiger 1996). This heat avoidance mechanism is effective for irrigated production in very hot environments, however, it is detrimental to WUE that is essential to obtain high productivity under limited water availability. Although the results of that study can not be extrapolated from *GB* to the other cultivated cotton species, some findings suggest that a similar mechanism acts also in *GH*. In hot environments stomatal behavior of upland cotton does not maximize WUE, unless the crop is water-stressed (Hutmacher and Krieg 1983; Radin 1989). Stomatal conductance in field-grown cotton (*GH*) is typically extremely high and exerts little influence over transpiration rate (Ackerson and Krieg 1977; Radin, Kimball, Hendrix and Mauney 1987). In conclusion, cotton breeding, traditionally emphasizing maximal productivity under optimal soil and water conditions, has resulted in fairly good adaptation to the high temperatures characterizing arid regions, but poor adaptation to water limitation also prevailing in these regions.

Despite the apparent drought resistance of wild cotton species, we are not aware of any prior efforts to explore whether they might contain genes useful in improving drought resistance of cultivated cotton. This may be due in part to the unusual properties of interspecific hybrids such as varying degrees of sterility noted above, as well as pronounced segregation distortions and a tendency to preserve parental haplotypes as a result of complex multilocus epistatic interactions (Jiang et al. 2000; Waghmare et al. 2005). Exotic cotton genotypes present limitations in breeding programs due to unfavorable traits such as photoperiodic response, low yield and poor quality. A 'conversion' program (McCarty and Jenkins 1992) only partly ameliorates these problems.

6 Genomic Dissection of Cotton Responses to Abiotic Stress

The development of stress resistant crops has been hindered by low heritability of complex traits such as yield, and by lack of knowledge of physiological parameters that reflect genetic potential for improved productivity under water deficit. Using genetic mapping to dissect the inheritance of different complex traits in the same population is a powerful means to distinguish common heredity from casual associations between such traits (cf. Paterson, Lander, Hewitt, Peterson, Lincoln and Tanksley 1988). In principle, this can permit a direct test of the role of specific physiological traits in genetic improvement of plant productivity under abiotic stresses, such as those imposed by arid conditions.

Genetic mapping has been used in several crops to identify quantitative trait loci (QTLs) responsible for improved productivity under arid conditions and for morpho-physiological traits associated with drought response (reviewed by Tuberosa and Salvi 2006). However, in only a few studies were productivity and physiological differences genetically-mapped in the same populations, enabling researchers to identify possible association between these traits. The identification of drought-related QTLs plays a central role in crop improvement through marker assisted selection. In addition, major stress-related QTLs can be considered for positional cloning, aiming to manipulate the target trait more directly by genetic engineering. Furthermore, cloning stress-related QTLs might substantially contribute toward a better understanding of the genetic and functional basis of plant responses to abiotic stress (reviewed by Salvi and Tuberosa 2005).

In cotton, two generations (F2 and F3) of an interspecific population (*GH* var H23 x *GB* var F177) were used for genetic mapping of the crop's productivity and related physiological traits under two irrigation regimes (Saranga et al. 2001, 2004). A total of 79 QTLs were detected for ten measures of plant productivity and physiological variables. Productivity of cotton in well watered versus water-limited conditions was largely accounted for by different QTLs, indicating that adaptation to both conditions can be combined into the same genotype (Saranga et al. 2001; Paterson, Saranga, Menz, Jiang and Wright 2002). QTL likelihood intervals for high seed cotton yield and low osmotic potential corresponded in three of four possible genomic regions, implicating leaf osmotic potential as a major component of improved cotton productivity under arid conditions (Saranga et al. 2004). Two of these three loci mapped to homeologous (corresponding) locations on the two sub-genomes of tetraploid cotton (chromosomes 6 and 25), suggesting that a particularly important role of one or more ancestral genes in that region may have been retained since the A-D genome divergence and polyploid formation. QTLs for $\delta^{13}\text{C}$ showed only incidental association with productivity, indicating that selection for WUE alone may reduce water consumption but is not expected to either increase or decrease productivity. The finding that the *GH* allele was

favorable at some loci and the *GB* allele at other loci illustrates how recombination of favorable alleles from each of these species may form novel genotypes that are better adapted to arid conditions than either of the parental species. Near-isogenic lines (NILs) for selected QTLs have been recently developed and show a considerable potential for improving drought adaptation (Levi, Paterson, Barak, Yakir, Wang, Chee and Saranga 2009). These NILs may offer a powerful tool for the identification of the underlying gene(s) and physiological mechanism(s), by separating these QTLs from one another and from their genetic background.

Newly developed genomic tools can further facilitate the genomic dissection of cotton adaptations to stress conditions. More than 3400 sequence-tagged DNA marker loci have been mapped in the leading reference populations (Reinisch, Dong, Brubaker, Stelly, Wendel and Paterson 1994; Rong et al. 2004; Rong, Bowers, Schulze, Waghmare, Rogers, Pierce, Zhang, Estill and Paterson 2005). Molecular dissection of complex traits is well-established (cf. Wang, Dong and Paterson 1995; Jiang, Wright, El-Zik and Paterson 1998; Wright, Thaxton, Paterson and El-Zik 1998, 1999; Lan, Cook and Paterson 1999; Jiang et al. 2000; Saranga et al. 2001, 2004; Paterson et al. 2002; Chee, Draye, Jiang, Decanini, Delmonte, Bredhauer, Smith and Paterson 2005a, b; Draye, Chee, Jiang, Decanini, Delmonte, Bredhauer, Smith and Paterson 2005; Rong et al. 2005), based largely on the reference map cited. These tools, fully sequenced and publicly available in Genbank (Rong et al. 2004), have also long been freely disseminated as clones and are in use by the worldwide cotton research community. A physical map is presently being constructed (under NSF support) based on a ~10x BAC library for the genome of *G. raimondii*, the source of the smaller and less-repetitive of the tetraploid cotton subgenomes (D), setting the stage for linking mapped DNA markers to BAC 'contigs' containing virtually any cotton gene. Additional BAC libraries are publicly available for *GH* cultivar 'Acala Maxxa' (8.3x coverage, average insert 137 kb); *GB* 'Pima S6' (5x coverage, 100 kb inserts); *G. arboreum* (A- genome) 'AKA8401' (6x coverage, 120 kb inserts); *G. longicalyx* (F- genome, 8.3x coverage with 4.3x from *HindIII* and 4x from *MboI*, 125 kb inserts), and other genotypes. More than 250,000 expressed-sequence tags are available in Genbank and are thought to represent at least 33,665 *Gossypium* genes (Udall, Swanson, Nettleton, Percifield and Wendel 2006), providing an archive of cloned cotton genes and their sequences. A first step toward cotton genome sequencing in being made in a US Dept of Energy Joint Genome Institute Community Sequencing Program (JGI-CSP) project of 500 Mb scale (0.6x genome coverage), toward design of an optimal strategy for complete sequencing of *G. raimondii*.

The scope of gene expression studies has been significantly altered by the development of microarray technology permitting simultaneous analysis of transcript levels for thousands of genes in different physiological states or different genotypes. This method is highly suitable to revealing coordinated gene expression patterns and to discovering metabolic pathways involved in

complex traits and phenotypes, such as drought resistance (e.g. Cushman and Bohnert 2000; Hazen, Wu and Kreps 2003; Ozturk, Talame, Deyholos, Michalowski, Galbraith, Gozukirmizi, Tuberosa and Bohnert 2002; Seki, Narusaka, Abe, Kasuga, Yamaguchi-Shinozaki, Carninci, Hayashizaki and Shinozaki 2001). A cotton oligonucleotide microarray (www.cottonevolution.info/microarray) have been developed. A second array is being put into use based on short microarray probes ("Affymetrix-style" ~25-mers) that exploit the sequence differences between homoeologous loci that have accumulated since speciation of the A and D parental genomes, enabling one to distinguish each member of duplicated gene pairs (and thus to analyze the effects of polyploidy on expression levels) for approximately 3000 genes (Udall et al. 2006).

Finally, the joining of two divergent genomes into a common polyploid nucleus is among the most important genetic mechanisms in plant evolution, and contributes substantially to the productivity of many crops (cf. Simmonds 1979). For example, the yield and quality of tetraploid cottons far exceed the levels of diploids (Jiang et al. 1998). However, little is known about how divergent genomes that have recently-united in a common nucleus function in concert to confer adaptation to environmental stresses. Complementarities between the polyploid subgenomes exhibited superior levels of drought resistance in *GH* and *GB* (Saranga et al. 2001), and additional valuable complementary genes are expected to exist in the unexplored wild tetraploid species.

7 Transgenic Approach to Improve Abiotic Stress Resistance in Cotton

Changes occurring rapidly at the mRNA and protein levels lead to the drought tolerant state (Ingram and Bartels 1996). Genes induced during stress are broadly classified into two groups: (i) genes thought to be involved in protecting stressed cells, and (ii) genes implicated in regulation of signal transduction and gene expression (Shinozaki and Yamaguchi-Shinozaki 1997; Bartels and Sunkar 2005). The manipulation of such genes by transgenic techniques offers excellent opportunities to gain better insight into the particular mechanisms or genes controlling plant responses to abiotic stresses as well as to improve plant stress resistance (Bohnert et al., 1995; Bajaj, Targolli, Liu, Ho and Wu 1999; Umezawa, Fujita, Fujita, Yamaguchi-Shinozaki and Shinozaki 2006). In the past, most transgenic work that was successful in commercial agriculture was directed towards crop resistance to biotic stresses or to technological properties (reviewed by Sonnewald 2003). Studies addressing plant resistance to abiotic stresses have been confined so far to experimental laboratory work and to single gene approaches, which have led to marginal stress improvement (Ramanjulu and Bartels 2002). However, recent

advances suggest that rapid progress will be possible in the near future, with large economical impact in many areas of the globe (reviewed by Chaves and Oliveira 2004).

Transgenic *Arabidopsis* plants overexpressing a single gene for stress induced transcription factor, *DREB1A*, that binds to a *cis*-acting dehydration responsive element (DRE), were reported to enhance drought, salinity and freezing tolerance (Kasuga, Liu, Miura, Yamaguchi-Shinozaki and Shinozaki 1999), suggesting a common mechanism associated with these stresses. The use of a strong constitutive promoter (35S) to drive expression of *DREB1A* resulted in severe growth retardation under normal growing conditions, whereas a stress inducible promoter (*rd29A*) gave rise to greater stress tolerance with minimal effects on plant growth. In cotton DRE-binding protein 1 (*GhDBP1*) was induced by salinity and drought and found to act as a transcriptional repressor for DRE-mediated gene expression, suggesting antagonistic mechanisms controlling DRE element-dependent transcription (Huang and Liu 2006).

Alleviation of oxidative damage by the use of various antioxidants and ROS scavengers can enhance plant resistance to abiotic stress. Transgenic cotton seedlings expressing the tobacco glutathione S-transferase (GST) Nt107, a gene that has been implicated in antioxidant activity, did not provide adequate protection against oxidative stress (Light, Mahan, Roxas and Allen 2005). Overexpression of the *Arabidopsis* gene GF14 λ , encoding a 14-3-3 protein that interacts with proteins involved in numerous metabolic processes, including antioxidant activity, demonstrated a "stay-green" phenotype and improved tolerance to moderate water stress (Yan, He, Wang, Mao, Holaday, Allen and Zhang 2004).

The accumulation of Na⁺ ions into vacuoles through the operation of a vascular Na⁺/H⁺ antiporter provides an efficient way to avert the deleterious effect of Na⁺ in the cytosol and maintain osmotic balance by using Na⁺ (and Cl⁻) accumulated in the vacuole to derive water into the cells (Glenn, Brown and Blumwald 1999). A cDNA clone isolated from cotton (*GH*), *GhNHX1*, was characterized as a tonoplast Na⁺/H⁺ antiporter (Wu, Yang, Meng and Zheng 2004). Salt stress and abscisic acid induced the expression of *GhNHX1* in cotton with a higher expression in salt-tolerant cultivars than in salt-sensitive ones. Transgenic tobacco plants overexpressing *GhNHX1* exhibited higher salt tolerance than wild-type plants. In another study, transgenic cotton plants overexpressing an *Arabidopsis* vacuolar Na⁺/H⁺ antiporter, *AtNHX1*, exhibited improved photosynthetic rate, nitrogen assimilation and productivity in the presence of 200 mM NaCl in greenhouse conditions, as compared with wild-type plants (He, Yan, Shen, Fu, Holaday, Auld, Blumwald and Zhang 2005). Moreover these transgenic plants produced more fiber with better quality also under field conditions, thus showing the applicability of the transgenic approach to improving crop performance under abiotic stress.

8 Synopsis

A large part of cotton's potential productivity is lost to various abiotic stresses. Since agricultural inputs are becoming more costly and scarce, improving stress resistance of crop plants is vital to meet future demands for food and fiber. Wild *Gossypium* species, native to semi-arid, semi-tropical environments, offer genetic diversity relevant to stress adaptation which is potentially available for exploitation in cotton improvement programs. The development of stress-tolerant crops by traditional breeding has been hindered by the low heritability of traits such as yield, particularly under stress, and by its large 'genotype x environment' interaction (Ali, Aslam, and Hussain 2005; Saranga et al. 2001). Modern genomic techniques have aided tremendously in identifying stress-related QTLs and diagnostic DNA markers in a number of crops including cotton. More recently, it has become possible to identify genes involved in plant response to stress and to transfer them into an alien background. These genomic tools promise to increase our knowledge of mechanisms conferring stress resistance in cotton and identify the underlying candidate genes and pathways. The development of stress resistant cultivars has become in recent years more feasible than ever, however, it is still a complex task that requires (and warrants) an extensive interdisciplinary physiological, genetic and genomic research effort.

Acknowledgment We thank Z. Peleg for his valuable comments on the manuscript.

References

- Ackerson, R.C. and Krieg, D.R. (1977) Stomatal and non-stomatal regulation of water use in cotton, corn and sorghum. *Plant Physiol.* 60, 850–853.
- Ali, Y., Aslam, Z., and Hussain, F. (2005) Genotype and environment interaction effect on yield of cotton under naturally salt stress condition. *Int. J. Environ. Sci. Tech.* 2, 169–173.
- Araus J.L., Slafer G.A., Reynolds M.P., Royo C. (2002) Plant breeding and drought in C_3 cereals: What should we breed for? *Ann. Bot.* 89, 925–940.
- Araus, J.L., Bort, J., Steduto, P., Villegas, D. and Royo, C. (2003) Breeding cereals for Mediterranean conditions, ecophysiology clues for biotechnology application. *Ann. Appl. Biol.* 142, 129–141.
- Ashraf, M. (2002) Salt tolerance of cotton, some new advances. *Critical Rev. Plant Sci.* 21, 1–30.
- Ashraf, M. and Ahmad, S. (2000) Influence of sodium chloride on ion accumulation, yield components and fibre characteristics in salt-tolerant and salt-sensitive lines of cotton (*Gossypium hirsutum* L.). *Field Crops Res.* 66, 115–127.
- Bajaj, S., Targolli, J., Liu, L.F., Ho, T.H.D. and Wu, R. (1999) Transgenic approaches to increase dehydration-stress tolerance in plants. *Mol. Breed.* 5, 493–503.
- Bartels, D., Furini, A., Ingram, J. and Salamini, F. (1996) Responses of plant to dehydration stress, A molecular analysis. *Plant Growth Reg.* 20, 111–118.
- Bartels, D. and Sunkar, R. (2005) Drought and Salt Tolerance in Plants. *Critical Rev. Plant Sci.* <http://www.informaworld.com/smpp/title~content=t713400911~db=all~tab=issueslist~branches=24-v2424>, 23–58.

- Blum A. (2005) Drought resistance, water-use efficiency, and yield potential – are they compatible, dissonant, or mutually exclusive? *Aust. J. Agric. Res.* 56, 1159–1168.
- Bohnert, H.J., Nelson, D.E. and Jensen, R.G. (1995) Adaptations to Environmental Stresses. *Plant Cell*, 7, 1099–1111.
- Boulanger, J. and Pinheiro, D. (1971) Evolution de la production au nord-est Brasil. *Cotton Fibr. Trop.* 26, 319–353.
- Boyer, J.S. (1982) Plant productivity and environment. *Science* 218, 443–448.
- Bray, E. (1993) Molecular responses to water deficit. *Plant Physiol.* 103, 1035–1040.
- Brubaker, C. L., Paterson, A. H., and Wendel, J. F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome*, 42, 184–203.
- Bruce, W.B., Edmeades, G.O. and Barker, T.C. (2002) Molecular and physiological approaches to maize improvement for drought tolerance. *J. Exp. Bot.* 53, 13–25.
- Chaves, M.M., Maroco, J.P. and Pereira, J.S. (2003) Understanding plant responses to drought – from genes to the whole plant. *Functional Plant Biol.* 30, 239–264.
- Chaves, M.M. and Oliveira, M.M. (2004) Mechanisms underlying plant resilience to water deficits prospects for water-saving agriculture. *J. Exp. Bot.* 55, 2365–2384.
- Chee, P., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, B., Smith, C.W. and Paterson, A.H. (2005a) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach, I. Fiber elongation. *Theor. Appl. Genet.* 111, 757–763.
- Chee, P., Draye, X., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, B., Smith, C.W. and Paterson, A.H. (2005b) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach, III. Fiber length. *Theor. Appl. Genet.* 111, 772–781.
- Condon, A.G. and Hall, A.E. (1997) Adaptation to diverse environments, Variation in water-use efficiency within crop species. In: L.E. Jackson (Ed.), *Ecology in Agriculture*. Academic Press, San-Diego, pp. 79–116.
- Condon, A.G., Richards, R.A., Rebetzke, G.J. and Farquhar, G.D. (2004) Breeding for high water-use efficiency. *J. Exp. Bot.* 55, 2447–2460.
- Cornish, K., Radin, J.W., Turcotte, E.L., Lu, Z. and Zeiger, E. (1991) Enhanced photosynthesis and stomatal conductance of Pima cotton (*Gossypium barbadense* L.) bred for increased yield. *Plant Physiol* 97, 484–489.
- Cowan, I.R. (1986) Economics of carbon fixation in higher plants. In: T.J. Givnish (Ed.), *On the Economy of Plant Form and Function*. Cambridge University Press, Cambridge, pp. 133–170.
- Cushman, J.C. and Bohnert, H.J. (2000) Genomic approaches to plant stress tolerance. *Curr. Opin. Plant Biol.* 3, 117–124.
- Draye, X., Chee, P., Jiang, C., Decanini, L., Delmonte, T., Bredhauer, B., Smith, C.W. and Paterson, A.H. (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach, II. Fiber fineness. *Theor. Appl. Genet.* 111, 764–771.
- Edmeades, G.O., Cooper, M., Lafitte, R., Zinselmeier, C., Ribaut, J.-M., Habben, J.E., Löffler, C. and Banziger, M. (2001) Abiotic stresses and staple crops. In: J. Nosberger, H.H. Geiger, and P.C. Struik (Eds.), *Crop Science, Progress and Prospects. Proceedings of the Third International Crops Science Congress*, 17–21 August, (2000). CABI, Wallingford, UK, pp. 137–154.
- Farquhar, G.D., O’Leary, M.H. and Berry, J.A. (1982) On relationship between carbon isotope discrimination and intercellular carbon dioxide concentration in leaves. *Aust. J. Plant Physiol.* 9, 121–137.
- Fryxell, P.A. (1979) *The Natural History of the Cotton Tribe*. Texas A&M Univ. Press, College Station, TX.
- Fryxell, P.A. (1984) Taxonomy and germplasm resources. In: R.J. Kohel and C.F. Lewis (Ed.), *Cotton. Agron. Monogr.* 24. ASA, CSSA, and SSSA, Madison, WI. pp. 27–57.

- Fryxell P.A. (1992) A revised taxonomic interpretation of *Gossypium* L. (*Malvaceae*). *Rhedeia*, 108–165.
- Glenn, E.P., Brown, J.J. and Blumwald, E. (1999) Salt Tolerance and Crop Potential of Halophytes. *Critical Rev. Plant Sci.* <http://www.informaworld.com/smpp/title~content=t713400911~db=all~tab=issueslist~branches=18-v1818>, 227–255.
- Gotemare, V. and Singh P. (2004) Use of wild species for cotton improvement in India. *ICAC Recorder*, Vol. XXII, 12–14.
- Gur, A., and Zamir, D. (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol.* 2, 1610–1615.
- Hamdy, A., Ragab, R. and Scarascia-Mugnozza, E. (2003) Coping with water scarcity, water saving and increasing water productivity. *Irrig. Drain.* 52, 3–20.
- Hazen, S.P., Wu, Y. and Kreps, J.A. (2003) Gene expression profiling of plant responses to abiotic stress. *Funct. Integ. Genomics.* 3, 105–111.
- He, C.X., Yan, J.Q., Shen, G.X., Fu, L.H., Holaday, A.S., Auld, D., Blumwald, E. and Zhang, H. (2005) Expression of an *Arabidopsis* vacuolar sodium/proton antiporter gene in cotton improves photosynthetic performance under salt conditions and increases fiber yield in the field. *Plant Cell Physiol.* 46, 1848–1854.
- Huang, B. and Liu, J.Y. (2006) A cotton dehydration responsive element binding protein functions as a transcriptional repressor of DRE-mediated gene expression. *Biochem. Biophys. Res. Commun.* 343, 1023–1031.
- Hutmacher, R.B. and Krieg, D.R. (1983) Photosynthetic rate control in cotton – stomatal and nonstomatal factors (*Gossypium hirsutum*). *Plant Physiol.* 73, 658–661.
- Ingram, J. and Bartels, D. (1996) The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 377–403.
- Jiang, C., Chee, P., Draye, X., Morrell, P., Smith, C.W. and Paterson, A.H. (2000) Multi-locus interactions restrict gene flow in advanced-generation interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54, 798–814.
- Jiang, C., Wright, R., El-Zik, K. and Paterson, A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Nat. Acad. Sci. USA*, 95, 4419–4424.
- Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat. Biotechnol.* 17, 287–291.
- Kohel, R.J. (1974) Influence of certain morphological characters on yield. *Cotton Grow. Rev.* 51, 281–292.
- Kramer, P.J. (1980). Drought, stress, and the origin of adaptation. In: N.C. Turner and P.J. Kramer (Eds), *Adaptation of Plants to Water and High Temperature Stress*. John Wiley and Sons, NY, pp. 7–20.
- Lan, T-H., Cook, C. and Paterson, A.H. (1999) Identification of a RAPD marker linked to a male-fertility restoration gene in cotton (*Gossypium hirsutum* L.). *J. Agr. Genomics*, Volume 4. <http://www.ncgr.org/ag/jag/papers99/paper299/indexp299.html>.
- Lee, J.A. (1984) Cotton as a world crop. In: R.J. Kohel and C.F. Lewis (Eds.), *Cotton*. Agron. Monogr. 24. ASA, CSSA, and SSSA, Madison, WI, pp. 6–25.
- Levitt, J. (1972) *Responses of Plant to Environment Stress*. Academic Press. N.Y.
- Levi, A., Paterson, A.H., Barak, V., Yakir, D., Wang, B., Chee, P.W. and Saranga, Y. (2009) Field evaluation of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. *Mol. Breeding* (in press).
- Light, G.G., Mahan, J.R., Roxas, V.P. and Allen, R.D. (2005) Transgenic cotton (*Gossypium hirsutum* L.) seedlings expressing a tobacco glutathione S-transferase fail to provide improved stress tolerance. *Planta* 222, 346–354.
- Lu, Z., Chen, J., Percy, R.G., Sharifi, M.R., Rundel, P.W. and Zeiger, E. (1996) Genetic variation in carbone isotope discrimination and its relation to stomatal conductance in Pima cotton (*Gossypium barbadense*). *Aust. J. Plant Physiol.* 23, 127–132.

- Maroco, J.P., Petreira, J.S. and Chaves, M.M. (2000) Growth, photosynthesis and water-use efficiency of two C₄ Sahelian grasses subjected to water deficits. *J. Arid. Env.* 45, 119–137.
- McCarty, J. and Jenkins, J. (1992) Cotton germplasm, Characteristics of 79 day-neutral primitive race accessions. *Miss. Agric. For. Expt. Stn., Tech. Bull.* 184.
- Meredith, W.R.Jr., and Bridge, R.R. (1984) Genetic contributions to yield changes in Upland cotton.. In: W.R. Fehr (Ed.), *Genetic contribution to yield gains of five major crop plants*. ASA and CSSA, Stoneville, MS, pp. 75–87.
- Mimura, T., Kura-Hotta, M., Tsujimura, T., Ohnishi, M., Miura, M., Okazaki, Y., Mimura, M., Maeshima, M. and Washitani-Nemoto, S. (2003) Rapid increase of vacuolar volume in response to salt stress. *Planta*, 216, 397–402
- Mooney, H.A., Percy, R.W. and Ehleringer, J. (1987) Plant physiology ecology today. *BioScience* 37, 18–20.
- Morgan, J.M. (1984) Osmoregulation and water stress in higher plants. *Annu. Rev. Plant. Physiol.* 35, 299–319.
- Munns, R. (2005) Genes and salt tolerance, bringing them together. *New Phytol.* 167, 645–663.
- Munns, R., James, R.A., Lauchli, A. (2006) Approaches to increase the salt tolerance of wheat and other cereals. *J. Exp. Bot.* 57, 1025–1043.
- Ozturk, N.Z., Talame, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R. and Bohnert, H.J. (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.* 48, 551–573.
- Parry, M.A.J., Flexas, J. and Medrano, H. (2005) Prospects for crop production under drought, research priorities and future directions. *Ann. Appl. Biol.*, 147, 211–226.
- Paterson, A. H., Saranga, Y., Menz, M., Jiang, C. and Wright, R. (2002) QTL Analysis of Genotype x Environment Interactions Affecting Cotton Fiber Quality. *Theor. Appl. Genet.*, 106, 384–396.
- Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. and Tanksley, S.D. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726.
- Pettigrew, W.T. (2004) Physiological consequences of moisture deficit stress in cotton. *Crop Sci.* 44, 1265–1272.
- Pickergill, B., Barrett, S.C.H. and de Andrade-Lima, D. (1975) Wild cotton in northeast Brazil. *Biotropica* 7, 42–54.
- Plucknett, D.L., Smith, N.J.H., Williams, J.T. and Anishetty, N.M. (1987) *Gene Banks and the World's Food*. Princeton University Press, Princeton, NJ, USA.
- Quisenberry, J.E., Jordan, W.R., Roark, B.A. and Fryrear, D.W. (1981) Exotic cottons as genetic sources for drought resistance. *Crop Sci.* 21, 889–895.
- Radin, J.W. (1989) When is stomatal control of water loss consistent with the thermal kinetic window concept? In: *Proc Beltwide Cotton Conf.*, Nat. Cotton Council Am., Memphis, TN, pp 46–49.
- Radin, J.W., Kimball, B.A., Hendrix, D.L. and Mauney, J.R. (1987) Photosynthesis of cotton plants exposed to elevated levels of carbon dioxide in the field. *Photosynth Res.* 12, 191–203.
- Radin, J.W., Lu, Z., Percy, R.G. and Zeiger, E. (1994) Genetic variability for stomatal conductance in Pima cotton and its relation to improvements of heat adaptation. *Proc. Nat. Acad. Sci. USA* 91, 7217–7221.
- Ramanjulu, S. and Bartels, D. (2002) Drought- and desiccation-induced modulation of gene expression in plants. *Plant, Cell and Environ.* 25, 141–151.
- Reinisch, A.R., Dong, J.-M., Brubaker, C., Stelly, D., Wendel, J. and Paterson, A.H. (1994) A detailed RFLP map of cotton (*Gossypium hirsutum* x *G. barbadense*), Chromosome organization and evolution in a disomic polyploid genome. *Genetics*, 138, 829–847.

- Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C.-H., Pierce, G.J., Rainey K.M., Rastogi, V.K., Schulze, S.R., Trolinder, N.L., Wendel, J.F., Wilkins, T.A., Williams-Coplin, T.D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. and Paterson, A.H. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166, 389–417.
- Rong, J., Bowers, J.E., Schulze, S.R., Waghmare, V., Rogers, C., Pierce, G., Zhang, H., Estill, J. C. and Paterson, A.H. (2005) Comparative genomics of *Gossypium* and *Arabidopsis*, Unraveling the consequences of both ancient and recent polyploidy. *Genome Res.* 15, 1198–1210.
- Rosenow, D.T., Quisenberry, J.E., Wendt, C.W. and Clark, L.E. (1983) Drought tolerant sorghum and cotton germplasm. *Agric. Water Manag.* 7, 207–222.
- Salvi, S. and Tuberosa, R. (2005) To clone or not to clone plant QTLs, present and future challenges. *Trends in Plant Sci.* 10, 297–304.
- Saranga, Y., Flash, Y. and Yakir, D. (1998) Variation in water-use efficiency and its relation to carbon isotope ratio in cotton. *Crop Sci.* 38, 782–787.
- Saranga, Y., Jiang, C., Wright, R., Yakir, D. and Paterson, A.H. (2004) Genetic dissection of cotton physiological responses to arid conditions and their inter-relationships with productivity. *Plant Cell Environ.* 27, 263–277.
- Saranga, Y., Menz, M., Jiang, C., Wright, R., Yakir, D. and Paterson, A.H. (2001) Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. *Genome Res.* 11, 1988–1995.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. and Shinozaki, K. (2001) Monitoring the expression pattern of 1300 arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13, 61–72.
- Shinozaki, K. and Yamaguchi-Shinozaki, K. (1997) Gene expression and signal transduction in water-stress response. *Plant Physiol.* 115, 327–334.
- Silva, T.M.S., Camara, C.A., Medeiros, F.D., Oliveira, E.J., Agra, M.F., Harley, R.M. and Giulietti, A.B. (2006) Phaeophytins from *Gossypium mustelinum* Miers ex Watt (Malvaceae). *Biochem. System. Ecol.* 34, 263–264.
- Simmonds, N.W. (1979) *Evolution of Crop Plants*. Longman Scientific and Technical Press, Essex, England.
- Singh, R.P., Prasad, P.V.V., Sunita, K., Giri, S.N. and Reddy, K.R. (2007) Influence of high temperature and breeding for heat tolerance in cotton, a review. *Adv. Agron.* 93, 313–385.
- Sonnewald, U. (2003) Plant biotechnology, from basic science to industrial applications. *J. Plant Physiol.* 160, 723–725.
- Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps, unlocking genetic potential from the wild. *Science* 277, 1063–1066.
- Tuberosa, R. and Salvi, S. (2006) Genomics-based approaches to improve drought tolerance of crops. *Trends in Plant Sci.* 11, 405–412.
- Turner, N.C. (1986) Crop water deficits, a decade of progress. *Adv. Agron.* 39, 1–51.
- Udall, J.A., Swanson, J.M., Nettleton, D., Percifield, R.J. and Wendel, J.F. (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173, 1823–1827.
- Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Engineering drought tolerance in plants, discovering and tailoring genes to unlock the future. *Curr. Opin. Biotechnol.* 17, 113–122.
- Vinocur, B. and Altman, A. (2005) Recent advances in engineering plant tolerance to abiotic stress, achievements and limitations. *Curr. Opin. Biotechnol.* 16, 123–132.
- Vorosmarty, C.J., Green, P., Salisbury, J. and Lammers, R. B. (2000) Global water resources, vulnerability from climate change and population growth. *Science* 289, 284–288.

- Waghmare, V.N., Rong, J., Rogers, C.J., Pierce, G.J., Wendel, J.F. and Paterson, A.H. (2005) Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *G. tomentosum*. *Theor. Appl. Genet.* 111, 665–676.
- Wendel, J. (1989) New World tetraploid cottons contain Old World cytoplasm. *Proc. Natl. Acad. Sci. USA* 86, 4132–4136.
- Wang, G., Dong, J. and Paterson, A.H. (1995) The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germplasm, molecular analysis of introgressive plant breeding. *Theor. Appl. Genet.*, 91, 1153–1161.
- Wright, R., Thaxton, P., Paterson, A.H. and El-Zik, K. (1998) Polyploid formation in *Gossypium* has created novel avenues for response to selection for disease resistance. *Genetics* 149, 1987–1996.
- Wright, R., Thaxton, P., Paterson, A.H. and El-Zik, K. (1999) Molecular mapping of genes affecting pubescence of cotton. *J. Heredity*, 90, 215–219.
- Wu, C.A., Yang, G.D., Meng, Q.W. and Zheng, C.C. (2004) The cotton *GhNHX1* gene encoding a novel putative tonoplast Na⁺/H⁺ antiporter plays an important role in salt stress. *Plant Cell Physiol.* 45, 600–607.
- Yakir, D., De Niro, M.J. and Ephrath, J.E. (1990) Effect of water stress on oxygen, hydrogen and carbon isotope ratios in two species of cotton plants. *Plant Cell Environ.* 13, 949–955.
- Yan, J.Q., He, C.X., Wang, J., Mao, Z.H., Holaday, S.A., Allen, R.D. and Zhang, H. (2004) Overexpression of the *Arabidopsis* 14–3-3 protein GF14 λ in cotton leads to a “Stay-Green” phenotype and improves stress tolerance under moderate drought conditions. *Plant Cell Physiol.* 45, 1007–1014.

Bridging Traditional and Molecular Genetics in Modifying Cottonseed Oil

Qing Liu, Surinder Singh, Kent Chapman, and Allan Green

Abstract Cotton breeding has traditionally focused on improving crop productivity and fiber quality. While this is still the major objective, there is increased interest in improving the nutritional and functional properties of the cotton seed oil. The nutritional and industrial value of cottonseed oil, like other vegetable oils, is determined by its fatty acid profile. Conventional genetics and breeding approaches to altering fatty acid composition have explored natural variation and induced mutations, but have had little if any impact on germ-plasm development. Recent advances in the understanding of the basic biochemistry of seed oil biosynthesis, coupled with identification of genes for oilseed modification, have set the stage for the genetic engineering of cottonseed to produce designer oils tailored for specific applications. Considerable progress has been achieved in altering the relative levels of the existing fatty acids in cottonseed oil for enhanced nutritional value and expanded industrial applications. Transgenic production in cottonseed oil of novel fatty acids with high industrial value can contribute to the replacement of non-renewable petroleum feedstocks with renewable and sustainable bio-based feedstocks. As functional genomics progresses further non-transgenic methods, such as TILLING, could facilitate the detection and selection of mutations not achievable through traditional breeding.

1 Introduction

Cotton is a dual purpose crop, producing both fiber and seed as valuable primary agricultural products. In the process of ginning the cotton boll, the fiber is separated from the seed and used in the textile industry. The separated seed at this point is called fuzzy cottonseed, and can either be further processed or be used directly as cattle feed. Fuzzy cottonseed is processed into four major

Q. Liu (✉)
CSIRO Plant Industry, ACT, 2601, Australia
e-mail: qing.liu@csiro.au

products including hull (26%), linters (9%), oil (16%) and meal (45%), with 4% lost in processing (Cherry and Leffler 1984). Cottonseed oil is the most valuable of these products.

Although cotton is primarily grown for its fiber, it is also the world's sixth largest source of vegetable oil. In terms of annual crushing, it is the world's third leading oilseed. A cotton plant will produce about 1.55 tons of seed for every ton of lint. Normally, cottonseed products represent approximately 15% of farm value of the cotton crop, making it a valuable co-product of cotton fiber production. Furthermore, the continued demand for cotton fiber by the textile industry ensures the production of a large supply of cottonseed.

Cottonseed oil has a long tradition of use in food processing, which can be traced back to ancient times when Asians in cotton growing regions developed crude methods for extracting oil from the seeds. However, the use of cottonseed oil on a commercial scale is of more recent origin, as modern crushing and refinery processes were developed or perfected in the United States. In fact, cottonseed oil was the first and dominant vegetable oil produced in the United States until the phenomenal growth of soybean in the 1940's. Similarly, cottonseed oil was the major food grade oil in Australia until 1950's when its dominant position was surpassed initially by sunflower and subsequently by canola.

Since cottonseed oil has a bland, neutral flavor that does not mask the inherent flavor of food, it is a popular and widely used oil for deep frying in the snack food and food service sector (Jones and King 1993). Cottonseed oil is also a key ingredient in many marinades, dressings, pastries, margarines, and shortenings. Currently in the US, cottonseed oil ranks third in volume behind soybean and corn oil representing about 5–6% of the total domestic fat and oil supply. The US is also the most significant cottonseed oil exporter in the world with an annual volume of about 100,000 metric tons. The price of cottonseed oil on the international market is similar to that of other commodity vegetable oils, such as soybean, corn, canola and sunflower. In these other major vegetable oilseed crops major breeding programs have placed considerable emphasis on improvements in nutritional value, functional properties and yield of oil. Continuous improvement in quality has intensified competition among oilseeds and has left cottonseed oil with significantly fewer customers and placed the industry in a vulnerable position. One reason for the lack of attention to oil improvement in cottonseed has been the increased focus on the value of whole cottonseed as a cost-effective feed for cattle. It was discovered that the tough seed-coat and fuzzy lint on the seeds could give the right amount of protection from hydrogenation in the rumen and provide a natural rumen bypass effect that improved milk output and milk fat content (Coppock, Lanham, and Horner 1987; Mohamed, Satter, Grummer, and Ehle 1988). Consequently since the 1980s, the feeding of whole cottonseed to cattle has risen steadily and caused a continual decline in the availability of cottonseed for crushing. Currently, less than half of the cottonseed produced in the

US goes for crushing, whereas this proportion was more than 90% in the early 1970s (Whillhit 2003).

Refined cottonseed oil contains more than 97% triglycerides which are neutral lipids, consisting of three fatty acids esterified to a glycerol backbone (O'Brien 2002). The nutritional and industrial value of cottonseed oil, like other vegetable oils, is determined by its composition of fatty acids with distinct carbon chain length and level of unsaturation. Cottonseed oil produced by traditional upland cotton (*Gossypium hirsutum*) typically consists of 26% palmitic acid, 2% stearic acid, 13% oleic acid and 58% linoleic acid (Table 1; Cherry 1983; O'Brien 2002). In addition to these common fatty acids which are also present in most other temperate zone oilseed crops, cottonseed oil also contains low levels (0.5–1%) of cyclopropenoid fatty acids (CPFAs), mainly malvalic and sterculic acids (Shenstone and Vickery 1961; Cherry 1983; Cherry, Kohel, Jones, and Powell 1986).

Table 1 Fatty acid composition of some edible vegetable oilseeds and genetic variants

Species	Fatty acid composition (%)*					Reference
	16:0	18:0	18:1	18:2	18:3	
Cotton	26	2	13	58	<1	
Soybean						
conventional	11	4	23	54	8	
high oleic	8	3	84	3	1	a
low palmitic	4	3	25	58	8	b
high stearic	9	26	18	39	8	c
high palmitic	25	4	16	44	10	b
low linolenic	12	6	39	41	2	b
Sunflower						
conventional	7	5	25	62	<1	
high oleic	3	2	90	4	<1	d
mid oleic	4	4	66	25	<1	d
high stearic	5	26	14	55	<1	e
Peanut						
conventional	13	3	38	45	<1	
high oleic	7	2	83	7	<1	f
Safflower						
conventional	7	3	18	71	<1	
high oleic	5	2	78	14	<1	g
Canola						
conventional	4	2	62	22	10	
high oleic	3	2	85	4	4	h
high stearic	4	35	41	14	6	i

*Minor fatty acids not included.

References: a, Kinney 1996a; b, Tarrago-Trani et al. 2006; c, Rahman et al. 1995; d, Kleingartner 2002; e, Osorio et al. 1995; f, Andersen et al. 1998; g, Fernández-Martínez et al. 1993; h, Möllers 2002; i, Knutzon et al. 1992.

A high level of saturated fatty acids, mainly palmitic acid contributes to the oxidative stability of cottonseed oil by offsetting the greater instability of the other unsaturated fatty acid components. It also imparts the high melting point required for making such products as margarine and shortening. On the other hand, palmitic is nutritionally undesirable because of its property of raising LDL-cholesterol, associated with increased risk of cardiovascular heart disease (CHD) (Lindsey, Benattar, Pronczuk, and Hayes 1990). In the US, all food products must be labeled for their total content of saturated fatty acids because of their impact on cardiovascular health. The American Heart Association recommends that the intake of saturated fat be limited to 7–10% or less of the total calories consumed each day. A food product cannot be advertised as “low in saturated fat” unless it contains 1 gram or less of saturated fat per serving (Lechtenstein, Appel, Brands, Carnethon, Daniels, Franch, Franklin, Kris-Etherton, Harris, Howard, Karanja, Lefevre, Rudel, Sacks, Van Horn, Winston, and Wylie-Rosett 2006).

Cottonseed oil also contains a high level of polyunsaturated fatty acid, with linoleic acid accounting for more than half of total fatty acids (Table 1). Albeit good for health, linoleic acid is oxidatively unstable and therefore limits the shelf life of the oil and makes it unsuitable for some food applications. One way to improve oxidative stability of vegetable oils involves partial hydrogenation, during which polyunsaturated fatty acids are transformed into more stable monounsaturated and saturated fatty acids. Because partial hydrogenation results in oils with an extended shelf life and greatly improved stability during frying, it was once regarded as a significant achievement in the food industry. Partial hydrogenation has also been used to convert cottonseed oil and some other vegetable oils into the higher melting point hard stocks needed for margarine and shortening manufacture, enabling the production of spreads that lacks cholesterol. However, there are a number of downsides to these oils or fats resulting from the partial hydrogenation process, including the increased processing cost, an undesirable flavor and distinctive taste in the food prepared with hydrogenated oils; and most importantly, the high level of *trans* fatty acids (TFA) produced by the hydrogenation process. TFA has been increasingly recognized to have significant LDL-cholesterol raising and HDL-cholesterol lowering properties and therefore increases the risk of cardiovascular disease based on evidence derived from epidemiologic and clinical studies (Oomen, Ocke, Feskens, van Erp-Baart, Kok, and Kromhout 2001; Maozaffarian, Katan, Ascherio, Stampfer, and Willett 2006). As a result an active campaign has been underway pursuing the removal of TFA from the food chain. In 2006 the US Food and Drug Administration (FDA) implemented a labeling requirement for TFA content in all food products, and similar regulations are expected to be implemented in many other Western countries in the near future. Alternative oils with similar functionality to partially hydrogenated oil but free of TFA are currently being actively sought.

Compared with polyunsaturated fatty acids, oleic acid is more stable towards oxidation both at ambient storage temperatures and at the high

temperatures which prevail during the cooking and frying of food. Therefore, oils with high amounts of oleic acid are considerably slower to develop rancidity during shelf-life or to oxidatively decompose during frying than those oils that contain high amounts of polyunsaturated fatty acids. Consequently raising of oleic acid at the expense of polyunsaturated fatty acids has been set as a primary goal in breeding programs concerning with oil quality in several oilseed crops (Howell 1971; Voelker and Kinney 2001).

In addition to oleic acid, stearic acid is also regarded as a desirable dietary component of seed oil. Stearic acid is a saturated fatty acid but it behaves differently from its shorter chain saturated fatty acid counterparts, such as palmitic acid and myristic acid, in that it does not raise LDL-cholesterol or total cholesterol (Bonanome and Grundy 1988; Zock and Katan 1992; Kris-Etherton, Mustad, and Derr 1993b). It is therefore regarded as a nutritionally benign saturate. Stearic acid has a high melting point and can impart the solidity and plasticity required for spread and shortening manufacture. Therefore, oils rich in stearic acid could be used as the hard stock for manufacturing of margarine and shortening (Tarrago-Trani, Philips, Lemar, and Holden 2006).

As well as these major components, cottonseed also contains a number of minor components that are of nutritional significance. On one hand it contains antinutritional components, such as CPFAs and gossypol, which impose significant limitations on use in animal feeds (Cherry 1983; O'Brien 2002). On the other hand, cottonseed oil also contains a number of lipid-soluble micronutrients, such as Vitamin E, phytosterols, and neuroactive *N*-Acylethanolamines that can have positive nutritional effects. Genetic alterations of these minor components in cottonseeds could add considerably to the value of cottonseed. Furthermore, new opportunities for cottonseed oil are emerging in industrial applications, such as in the production of biodiesel and chemical feedstocks, due to the increasing interest in using renewable plant oils as replacements for diminishing petroleum resources in the fuel and chemical sectors. Consequently, the cottonseed industry is facing both challenges and opportunities to meet the specific requirements of various food and industrial applications in a global market that is dominated by soybean and palm oils. Specific tailoring of cottonseed oil composition to match this range of diversified end uses has been limited using traditional genetics and breeding approaches, but is now being readily achieved through the use of recently developed molecular genetics and genetic engineering techniques.

2 Traditional Genetics and Breeding

Because cotton breeding has traditionally been dominated by considerations of fiber yield and quality, research on cottonseed oil composition has been a relatively neglected area. Prospects of increased utilization of cottonseed oil

as food, feed and biodiesel, together with intensified competition from improved forms of other oilseeds, has increased awareness of the opportunities to improve cottonseed oil quality. Improved seed oil content and fatty acid composition have become important considerations in cotton breeding programs in a number of countries, provided that they can be achieved without compromising fiber quality and yield.

Research on modifying fatty acid composition in cottonseed oil was intended to follow the examples from other oilseeds, which relied on naturally occurring variation resulting from spontaneous mutations. For example, through the use of a mass-screening method, a number of spontaneous mutants with reduced level of nutritionally undesirable erucic acid were found in rapeseed (Downey and Craig 1964). Canola quality oil was developed by combining some of these spontaneous mutants to give a very low level of erucic acid. Recent molecular studies revealed that these mutations were in two genes encoding isoforms of β -ketoacyl-CoA synthase, a key enzyme component of the fatty acid elongase complex that catalyses the conversion of oleic acid to erucic acid (Fourmann, Barrel, Renard, Pelletier, Delourme, and Brunel 1998). Other examples of successful traditional breeding for improvement of oil quality using spontaneous mutation include the production of high-stearic cultivars of soybean (Graef, Fehr, and Hamond 1985) and sunflower (Osorio, Fernandez-Martínez, Mancha, and Garcés 1995). However, spontaneous mutations occur at low frequency and they can be masked by normal alleles at homologous loci in polyploid plants. Increased frequency of mutations can be achieved through chemical or radiation induced mutagenesis, and this technique has been used effectively to alter fatty acid composition. One typical example is the genetic removal of the oxidatively unstable α -linolenic acid from linseed. A large number of seeds were treated with ethyl methane sulfonate (EMS) and individual mutants with reduced level of α -linolenic acid were identified (Green and Marshall 1984). A line with α -linolenic acid level as low as 2% of total fatty acids in seed oil was selected from the populations generated by crossing some of these single mutants giving rise to LinolaTM (Green and Dribnenki 1994).

The National Cotton Council in the USA has set a goal for incremental increase of cottonseed oil by 2 percentage points and there has been some progress. The breeding program of the Acala SJ series in California (Cherry, Kohel, Jones, and Powell 1981; Cherry 1983) increased content of oil in cottonseed from 19.0% to 21.8%, reduced palmitic acid from 23.3 to 22.7%, increased oleic acid from 16.6% to 17.3% and reduced cyclopropanoid fatty acids (CPFAs) from 0.9% to 0.8. However, compared to achievements in other oilseed crops, these changes are only modest and may reflect the narrow genetic base being used to improve cottonseed quality (see Brubaker, Lubbers and Percy Chapters of this volume).

Considerable resources have gone into collecting, developing, and maintaining cotton germplasm resources, and in evaluating some of them for seed oil quality (Percival and Kohel 1990; Lukonge, Kabuschnage, and Hugo 2007). Genetic variability for seed oil content in some wild *Gossypium* species has

been investigated. Seed oil content varied from 10.3 to 22.9% in 22 species of *Gossypium* and six races of *G. arboreum* (Gotmare, Singh, Mayee, Deshpande, and Bhadat 2004). However, the incompatibility, poor fiber quality and agronomic traits of the wild *Gossypium* relatives made it impossible to transfer the increased oil content trait to cultivated cotton. Furthermore, the elevation of oil content in some other crops was mainly achieved by combining several favorable alleles in one genetic background through many years of breeding. Chemical mutagenesis has resulted in a number of cotton mutants with increased seed oil content, and a mutant with reduced fuzzy lint on the surface of a cottonseed, which could lead to more efficient oil extraction (Auld, Bechere, Davis, Seip, and Brown 2006). It was envisaged that more than a 20% oil increase could be realized by combining these two types of mutations, without compromising fiber yield and quality. There are believed to be around 30 enzymatic reactions involved in converting acetyl-CoA generated during photosynthesis into seed storage oil triglycerides (see below), which makes it difficult to achieve significant improvements through conventional breeding approaches. Thus, despite many years of interest in improving cottonseed quality, breeding research aimed at altering the composition of cottonseed has had little, if any, impact on the improvement of cotton cultivars.

3 Molecular Genetics and Genetic Engineering

Genetic engineering provides a rapid and direct method for manipulating fatty acid composition in oilseeds, including cottonseed. Recent advances in understanding of the biochemical, cellular and molecular mechanisms of seed oil biogenesis, coupled with the cloning of many of the genes involved in this process, have facilitated the production of designer cottonseed oil with improved nutritional benefits and enhanced functional properties. Genetic engineering can also enable the production of high-value novel fatty acids in cottonseed oil to help meet the increasing demand for renewable energy and biodegradable raw materials.

During the late 1980s and 1990s, cotton was at the forefront of transgenic crops in commercial production. The first wave of transgenic cotton was made to ward off pests and to tolerate herbicides used to kill weeds, and offered advantages to farmers in the production phase of agriculture (see Davis Chapter, this volume). In contrast to these input-focused traits, the second wave of transgenic oilseed plants emerged with novel fatty acid composition that meet the needs of consumers and industry. Laurate canola (Voelker, Worrell, Anderson, Bleibaum, Fan, Hawkins, Radke, and Davies 1992; Voelker, Hayes, Cranmer, Turner, and Davies 1996) and high-oleic soybean (Kinney 1996a) are two specialty GM crops that have been deregulated in the US, although they are still far from becoming major commercial successes. Progress has been made in producing transgenic cotton rich in oleic and stearic

acids, and low in palmitic acid and CPFAs in seed oil (Chapman, Austin-Brown, Sparace, Kinney, Ripp, Pirtle, and Pirtle 2001; Liu, Singh, and Green 2002; Liu, Bao, Singh, Pons, Hurlstone, Dowd, McFadden, Pollard, Ohlrogge, and Green 2004; Sunikumar, Campbell, Puckhaber, Stipanovic, and Rathore 2005). The third wave of genetic modification of oilseeds will focus on the production of high value industrial fatty acids, with promising results already having been reported for the synthesis of epoxy-fatty acids in cottonseed oil (Zhou, Singh, Liu, and Green 2006).

Genetic modification of cottonseed oil has also been made more efficient through a series of methodological advancements in transgene expression system, plant regeneration from tissue culture and gene transformation *via Agrobacterium tumefaciens* or particle bombardment. The chimeric transgene typically contains a cloned gene of interest driven by a seed-specific promoter. Antibiotic or herbicide resistance is also included on the gene construct to enable a selection of successfully transformed embryogenic cells. Expression of a transgene or down-regulation of endogenous genes, or a combination thereof, can be used as a means to alter metabolic pathways in oil biosynthesis in cottonseed oil.

Down-regulation of endogenous gene expression in cotton can be achieved by several strategies, including antisense, co-suppression, ribozyme, or RNAi, all of which involve the introduction into the cotton cell of all or part of a gene homologous to the endogenous gene targeted for silencing (Knutzon, Thompson, Radke, Johnson, Knauf, and Kridl 1992; Kinney 1996a; Buhr, Sato, Ebrahim, Xing, Zhou, Mathiesen, Schweiger, Kinney, Staswick, and Clemente 2002; Waterhouse, Graham, and Wang 1998). Since the fatty acid composition of membrane lipids influences many aspects of development, as well as adaptation to abiotic conditions, it is crucial to restrict the modification to the lipids of the seed. A series of seed-specific promoters derived from the genes highly expressed in the seeds of dicot plants, such as napin, lectin, phaseolin and α -globulin genes have been successfully used in modifying cotton fatty acid composition (Chapman et al. 2001; Liu et al. 2002; Zhou et al. 2006; Sunilkumar, Campbell, Puckhaber, Stipanovic, and Rathore 2006). EST and microarray technologies have been used to expand the range of promoters that meet the quantitative and seed-specific gene expression requirements for modifying fatty acid composition or other seed constituents in seed oil (Girke, Todd, Ruuska, White, Benning, and Ohlrogge 2000). ESTs derived from seed cDNA libraries can be used to identify genes that are highly expressed in seeds. Microarray analysis can then provide information on which of these genes are expressed specifically in seeds, leading to the isolation of additional seed-specific promoters.

A widely recognized model of RNAi gene silencing is that the transgene with inverted repeat configuration generates double stranded RNA (dsRNA) which is cleaved into small interference RNA (siRNAs) of 21 to 26 nucleotides by an RNase termed Dicer (Bernstein, Caudy, Hammond, and Hannon 2001). The siRNAs are then incorporated into an RNA-induced silencing complex (RISC)

to be associated with the target mRNAs (Hammond, Bernstein, Beach, and Hannon 2000), and the activated RISC degrades the target mRNAs and leads to specific down-regulation of the target gene.

3.1 Oil Biosynthesis

Cottonseed oil is composed of triacylglycerols (TAGs) that are synthesized and deposited during seed development. The biochemical processes involved in biosynthesis of seed storage oils are well established, and involve metabolic pathways located in several different subcellular organelles (see reviews by Browse and Ohlrogge 1995; Ohlrogge and Jaworski 1997; Voelker and Kinney 2001).

The *de novo* biosynthesis of fatty acids occurs in the stroma of plastids in the developing seeds. Fatty acids are then exported in the form of acyl-CoA thioesters to the cytoplasmic endomembrane systems where modification of fatty acids occurs while attached to phospholipids. This is followed by TAG assembly and storage in the oleosomes. Fig. 1 represents a simplified schematic of fatty acid biosynthesis in developing cottonseed.

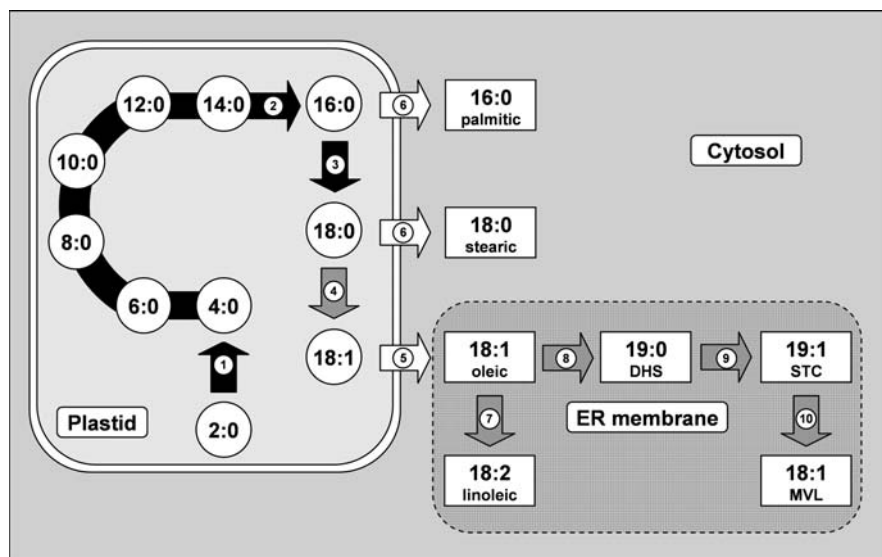


Fig. 1 A simplified schematic of fatty acid biosynthesis in developing cottonseed showing important enzymatic steps: 1. keto-acyl synthase III (KASIII), 2. keto-acyl synthase I (KASI), 3. keto-acyl synthase II (KASII), 4. Δ^9 -stearoyl-ACP desaturase (SAD), 5. oleoyl-ACP thioesterase, 6. acyl-ACP thioesterase, 7. Δ^{12} -oleoyl-lipid desaturase (FAD2), 8. cyclopropane fatty acid synthase (CPA-FAS), 9. cyclopropane fatty acid desaturase (CPA-DES), 10. α -oxidase. DHS: dihydrosterculic acid; STC, sterculic acid; MVL, malvalic acid.

In the plastids, the biotin-containing enzyme acetyl-CoA carboxylase (ACCase) catalyses the first committed step in the pathway by activating acetyl-CoA to the three carbon intermediate, malonyl-CoA, by addition of a carboxyl group. The malonyl group is then transferred from CoA to an acyl-carrier protein (ACP), which serves as the carrier for the growing fatty acid chain. Malonyl-ACP is reacted with a second acetyl-CoA condensing enzyme, ketoacyl-ACP synthase III (KASIII), resulting in a four carbon chain. The repeated process of adding two-carbon units on the elongated fatty acid chain is catalyzed by KASI leading to the formation of palmitoyl-ACP. KASII catalyzes the elongation of palmitoyl-ACP to stearoyl-ACP. A soluble stearoyl-ACP $\Delta 9$ -desaturase introduces the first double bond into stearoyl-ACP to convert it to oleoyl-ACP in the plastid. The growing saturated fatty acyl chain and the monounsaturated oleate are cleaved off the ACP by a specific thioesterase enzyme FatB or FatA, enabling them to exit the plastid into the cytoplasm. Saturated fatty acids released into the cytoplasm are not further modified. However, oleic acid can be further modified on the endoplasmic reticulum (ER) membranes by the action of membrane-bound desaturases. Phosphatidylcholine (PC) serves as a substrate for ER lipid modifying enzymes, such as fatty acid desaturase 2 (FAD2) which introduces a double bond into oleic acid on the sn-2 position of PC. Oleic acid is also the precursor for a group of cyclic fatty acids, cyclopropane and cyclopropenoid fatty acids (Fig. 1). All the modified and unmodified fatty acyl groups would then form a pool while attached to CoA. The fatty acyl groups are incorporated into membrane and storage lipids *via* the Kennedy pathway by the sequential esterification of glycerol-3-phosphate by the action of glycerol-3-phosphate acyltransferase (GPAT) to form lysophosphatidic acid, followed by lysophosphatidic acid acyltransferase (LPAAT) to form phosphatidic acid (PA). Dephosphorylation of PA by phosphatidic acid phosphatase results in the formation of diacylglycerol (DAG), which in developing seeds is then acylated to form triacylglycerol (TAG) by diacylglycerol acyltransferase (DGAT) (Browse and Ohlrogge 1995). TAG can also be formed in plants *via* two different acyl-CoA independent pathways, catalysed by phospholipid: diacylglycerol acyltransferase (PDAT) and diacylglycerol transacylase (DGTA) (Athenstaedt and Daum 2006).

3.2 Development of High-Oleic Cottonseed Oil (HO-CSO)

Because cottonseed oil normally contains relatively high levels of linoleic acid (57%) and low levels of oleic acid (15%), it implies that an active FAD2 converts about 80% of oleate formed to linoleate. Four different cDNAs encoding FAD2 have been isolated from upland cotton (Liu, Singh, Brubaker, Sharp, Green, and Marshall 1999b; Liu, Singh, Brubaker, and Green 1999a; Pirtle, Kongcharoensuntorn, Nampaisansuk, Knesek, Chapman, and Pirtle 2001;

Zhang, Pirtle, Park, Nampaisansuk, Neogi, Wanjie, Pirtle, and Chapman 2009), among which *ghFAD2-1* was determined to play a major role in the production of linoleic acid in cottonseed oil. Molecular analysis of gene expression suggested that the *ghFAD2-1* is specifically expressed in the developing seeds and its expression reaches the highest level at the middle maturity stage of seed development (Liu et al. 1999b).

A number of methods of posttranscriptional gene silencing (PTGS) have been used in down-regulating the expression of *ghFAD2-1* in a seed-specific manner in order to raise the oleic acid level at the expense of linoleic acid. Chapman et al. (2001) transformed cotton plants with a non-functional mutant allele of rapeseed *fad2* driven by a seed-specific phaseolin promoter in sense orientation. About half of the 43 transgenic lines that were generated showed moderately increased oleic acid content, ranging from 20–30%. In subsequent generations, lines with higher levels of oleic acid up to 40% were identified (Table 2). Interestingly, coincident with the increased percentage of oleic acid in these lines, it was observed that total oil content was reduced by 40% or more and reductions in total seed proteins also were observed (Chapman, Neogi, Hake, Stawska, Speed, Cotter, Garrett, Kerby, Richardson, Ayre, Ghosh, and Kinney 2008). This alteration was accompanied by fewer numbers of lipid bodies and protein bodies in cells of the transgenic seeds. The severely affected lines also showed a reduction in overall cotyledon thickness and a disruption in cellular and subcellular organization. One explanation is the non-functional rapeseed *fad2* allele was expressed, and the catalytically-inactive protein product had a dominant-negative mutation effect which resulted in detrimental effects on the normal physiology of seed development. It is believed by some that a dimer structure is required for FAD2 enzyme functionality, and it is tempting to speculate that FAD2 dimers incorporating the non-functional FAD2 enzyme could be embedded in the ER and impact both the desaturation and the normal function of lipid modifying enzymes in the synthesis of storage lipids in these cottonseed cells. Alternatively a more general interference with ER machinery by the *fad2* mutant gene product might explain a poor accumulation of oil and protein bodies. Regardless, the elevation of oleic acid percentage in cottonseeds by this *fad2*-mutant approach seems to occur by a

Table 2 Fatty acid composition of genetically modified cottonseed oils (CSO)

	Fatty acid composition (%) [*]					Reference
	16:0	18:0	18:1	18:2	18:1e [#]	
Conventional CSO	26	2	13	58	—	
High-stearic (HS-CSO)	17	38	10	38	—	a
Mid High-oleic (HO-CSO)	25	2	40	33	—	b
High-oleic (HO-CSO)	17	1	78	4	—	a
Vernolic CSO	22	2	27	32	17	c

^{*}Minor fatty acids not included. [#]18:1e: vernolic acid.

References: a, Liu et al. 2002; b, Chapman et al. 2001; c, Zhou et al. 2006.

mechanism different from RNA-based approaches (below) since neither RNAi nor antisense suppression of *FAD2* were reported to reduce oil content.

The expression of either antisense or hairpin (hp) structure (RNAi) of *ghFAD2-1* was also carried out in cotton, under the transcriptional control of a seed-specific promoter derived from soybean lectin gene (Liu et al. 2002). Both approaches were able to raise oleic acid level at the expense of linoleic acid, with RNAi being more effective. Two independent high-oleic lines with the maximum level of oleic acid (78%, Table 2), each carrying a single insertion of the transgene, were selected and bred to homozygosity. In contrast to the unfavorable effects on seed viability and cellular membranes of transforming cotton with mutated rapeseed *FAD2*, altered fatty acid composition in the seed oil was the only detectable difference between the *FAD2-RNAi* lines and a conventional cotton cultivar, Coker 315. The two selected *FAD2-RNAi* lines were further evaluated in a field trial in Narrabri, Australia in 2003/04 (Australian Government OGTR field trial DIR039/2003). Assessment of a range of agronomic and physiological performances including stand establishment, agronomic performance, flowering time, fiber yield, fiber quality, seed yield, oil content, fatty acid composition and germination rate in comparison to the Coker 315 control revealed that the high oleic trait did not negatively impact any aspect of plant performance (Liu et al. unpublished data).

Oxidative stability of the high-oleic cottonseed oil generated through RNAi-*ghFAD2-1* approach was assessed by measuring the development of aldehydes under accelerated oxidation conditions (Wijesundera et al. unpublished data). The total aldehyde concentration in the high-oleic oil was approximately 10-fold lower than that in the unmodified oil showing that the high-oleic cottonseed oil was significantly more resistant to oxidative deterioration than conventional cottonseed oil. Although no frying test has been performed with the high-oleic cottonseed oil, a superior performance conferred by the high oxidative stability is anticipated. The nutritional benefits of high-oleic cottonseed oil are not limited to the zero-trans fatty acids. It also contains significantly lowered palmitic acid. In addition to health benefits to consumers, the improved oxidative stability of high-oleic cottonseed oil could also deliver other benefits in industrial-scale frying in terms of reduced oil usage and prolonged product shelf-life (Kristott 2003).

The experience of the food industry has demonstrated that the profile of an optimal vegetable oil for frying may contain 65–75% oleic acid (Burton, Miller, Vick, Scarth, and Holbrook 2004). The presence of a moderately low level of linoleic acid is likely to provide a good balance between desirable flavor intensity, storage ability of fried food, and extended frying life of the high-oleic oil. The high-oleic cottonseed oil derived from the RNAi-*ghFAD2-1* lines contains only about 5% linoleic acid, and its effect on the flavor quality of fried food remains to be evaluated. There was sufficient variability between the RNAi-*ghFAD2-1* lines to enable selection of high-oleic lines with greater than 5% linoleic acid if that is eventually considered necessary to achieve optimal flavor performance.

Beyond the food industry, high-oleic cottonseed oil is also a potentially valuable feedstock for oleochemical applications. For such applications, an increase of oleic acid to over 90% would be of considerable advantage in enabling the production of higher quality chemicals with more uniform properties, such as has been achieved with high-oleic soybean oil that has already been commercialized by DuPont for lubricant formulations in the US (Cahoon 2003).

The most common industrial use for a high-oleic cottonseed oil would be as a desirable feedstock for biodiesel due to its high oxidative stability. Cottonseed oil derived biodiesel is commonly generated by oil transesterification in the presence of a catalyst. The derived methyl or ethyl esters can be used in an unmodified diesel engine in neat form or blended with petroleum diesel. Biodiesel performance of the current commodity cottonseed oil rich in linoleic acid is usually limited by the formation of gums and sediment in the fuel which can clog the fuel filter. However the high-oleic oil would have a high cetane value, high oxidative stability, high viscosity and excellent cold flow property due to low melting point (Kinney and Clemente 2005). The high oxidative stability of high-oleic cottonseed oil could also make it a suitable feedstock for biodegradable lubricants.

Apart from being used as biodiesel, oleoyl esters derived from high-oleic cottonseed oil could also be used as biodegradable replacements for mineral oils, such as hydraulic oils and lubricants etc. Further, the $\Delta 9$ position of the double bond in oleic acid offers a site for structural modifications in the oleochemistry processes. Purified oleic acid could be used to manufacture a range of oleochemicals like diacids, TMP-trioleate, ethylhexyloleate, glycerolmonooleate, decyloleate or oleyloleate which are found in the formulations of many hundreds of different oleochemical products (Hill 2000). For example, oleochemical diacids derived from chemically epoxidised oleic acid could be used to manufacture biodegradable polymers, such as polyurethanes, offering an interesting suite of properties including elasticity, flexibility, impact strength and hydrolytic stability (Hill 2000; Crandall 2002; Cahoon 2003). Currently the oleic acid used in oleochemistry is mainly derived from tallow or palm. Cottonseed oil with extra high levels of oleic acid could offer a sustainable and economical alternative and meet the ever increasing demand for such a value-added raw material.

3.3 High Stearic Cottonseed Oil (HS-CSO)

Specialty cottonseed oils having high contents of stearic acid are under development and could have valuable food and industrial applications. It is considered that vegetable oil with approximately 30% total saturates can make a suitable *trans*-free margarine through the process of transesterification (List, Steidley, and Neff 2000). If this could be achieved through use of high-stearic

oils rather than high-palmitic oils it would result in nutritionally superior products (Zock and Katan 1992; Nicolosi and Rogers 1997). Beyond food applications, stearic acid is also a valuable component in cosmetic products as an emulsifying agent or as an agent to achieve good consistency. Currently, the main source of stearic acid for the cosmetic industry is animal tallow. Cottonseed oil engineered to have a high content of stearic acid could provide a preferred alternative source because it would avoid the risk of bovine spongiform encephalopathy (BSE) and scrapie associated with animal products (Töpfer, Martini, and Schell 1995; Stoll, Lühs, Zarhloul, and Friedt 2005). Furthermore, cottonseed oil rich in stearic acid could also be used as biodiesel suitable for use in warm climates where it would remain liquid and have a high cetane value (Dyer and Mullen 2005).

In cotton, almost all stearic acid formed during seed development is subsequently desaturated to oleic acid by the action of stearoyl-ACP $\Delta 9$ -desaturase (SAD) with only around 2% remaining in the seed oil at maturity. SAD is encoded by a gene family consisting of at least five genes per diploid genome, as indicated by Southern blot analysis. cDNAs for each of these five *ghSAD* genes have been isolated from developing cottonseeds (Liu, Singh, Sharp, Green, and Marshall 1996; Liu et al. unpublished data). Primary data indicated that the *ghSAD-1* gene is expressed at significantly higher levels than the other four genes in developing seeds, but none of the five *SAD* genes are expressed exclusively in developing seeds. In other species, such as *Brassica napus* (Slocombe, Pianelli, Fairbairn, Bowra, Hatzopoulos, Tsiantis, and Murphy 1994) and flax (Fofana, Duguid, and Cloutier 2004; Fofana, Cloutier, Duguid, Ching, and Rampitsch 2006), the *SAD* genes are also expressed in other tissues such as ovaries, tapetum and pollen grains.

Transgenic cotton plants with increased stearic acid content in the seed oil have been generated by down-regulating SAD activity using either antisense or RNAi technology targeted against the *ghSAD-1* (Liu et al. 2002). A range of stearic acid levels was observed between transgenic lines as well as among individual seeds in a single transgenic line, with the highest level being 38% of total fatty acids (Table 2). This increase in stearic acid was accompanied by decreases in all other major fatty acids. Linoleic acid was reduced to 38% from the normal 56%; palmitic acid to 17% from 26% and oleic acid to 10% from 15%.

The high-stearic cottonseed oil has a remarkably increased melting point compared to the conventional cottonseed oil control. It is semi-solid at room temperature and has a sharp melting point at 15°C (Wijesundera et al. unpublished data). It appears that the high-stearic cottonseed oil contained insufficient stearic acid to provide the satisfactory texture and spreadability for direct use as solid fat unless used in combination with another appropriate fat.

An antisense strategy targeting the *SAD* gene has previously been successfully used in raising stearic acid content of rapeseed by up to 40% (Knutzon et al. 1992). Stearic acid has also been raised through modifications involving other genes. For example, transgenic expression of a *FatA* thioesterase gene

from mangosteen (*Garcinia mangostana*) in canola increased stearic acid from 2 to 22% (Hawkins and Kridl 1998), and overexpression of KASII in transgenic soybean resulted in a slight increase in stearic acid content (Kinney 1996b).

In contrast with high-oleic genotypes, the development of high-stearic cotton has been relatively slow, because of the reduced seed germination rate and compromised physiological performance associated with the high-stearic trait in modified cotton plants (Liu et al. 2002). This experience is similar to previous attempts to develop high-stearic canola using antisense against the *SAD* gene, where seeds with 40% stearic acid were produced but exhibited poor germination (Knutzon et al. 1992). Interestingly, this problem was not observed in canola with elevated stearic acid content generated by overexpression of the mangosteen *FatA*, however the level of stearic acid was significantly lower in that case.

The 40% stearic acid level in canola seed oil was shown to be associated with the significant elevation of stearic acid and corresponding reduction of unsaturated fatty acids in cell membranes, resulting in reduced membrane fluidity and impairment of membrane function that affected the development of the entire plant (Thompson and Li 1997). Cellular membranes are made up of a lipid bilayer into which are embedded various proteins that mediate such processes as transport, respiration, photosynthesis and signal transduction. The fatty acid composition of the membrane lipid is critically important for the cell's biochemical metabolism and environmental adaptation (Browse and Ohlrogge 1995; Ohlrogge and Jaworski 1997). Substantial reduction of unsaturated fatty acid in the high-stearic cottonseed might have led to serious membrane disruption, and hence impaired cellular function. Thus, the key impediment to developing viable high-stearic oilseeds may not lie in the altered composition of the triglycerides *per se*, but in the associated changes in membrane lipid composition.

3.4 Combined Increases in Oleic Acid and Stearic Acid

The ability to independently control levels of oleic and stearic acids through RNAi silencing of *Fad2* and *SAD-1* genes opens up the possibility to produce specific designer fatty acid compositions through combined and controlled silencing of both genes. For example, the combination of the high-oleic (HO) and high-stearic (HS) events described above resulted in simultaneous elevation of stearic and oleic acids in cottonseed oil (Liu et al. 2002). The HO/HS cottonseed oil contained 17% palmitic acid, 28% stearic acid, 50% oleic acid and 5% linoleic acid. Analysis of solid fat content (SFC) indicated that the HO/HS cottonseed oil possesses a sharp melting point around 20°C (Wijesundera et al. unpublished data). The combination of these traits therefore opens up the possibility of developing cottonseed oils that have specific solid fat properties without the need for hydrogenation. By choosing transgenic events with specific

intermediate levels of *FAD2* and *SAD-1* silencing it should be possible to precisely control the ratios of stearic, oleic and linoleic acids in cottonseed oils to meet desirable end product specifications. Such products could expand the utilization of cottonseed oil into new areas of the fats and oils industry, such as in the shortenings and cocoa-butter substitute segments.

3.5 Prospects for Further Modifications to Cottonseed Oil Composition

Recent developments in genetic engineering of fatty acid composition in seed oils show how amenable oil quality is to manipulation by molecular genetic approaches, and provide an indication of modifications that might be achievable in cottonseed oil in the future.

3.5.1 Altered Levels of Palmitic Acid

Except for palm oil, cottonseed contains the highest palmitic acid (26%) among the major commodity vegetable oils. Palmitic acid and some other shorter chain saturated fatty acids are widely reported to raise total plasma cholesterol and low density lipoprotein cholesterol levels (Kris-Etherton, Derr, Mitchell, Mustard, Russell, McDonnell, Salabsky, and Pearson 1993a). Significant reductions in palmitic acid would therefore increase the health appeal of cottonseed oil, and make it more competitive with nutritionally improved oils from other oilseeds. Reduction of palmitic acid could also eliminate the cloudiness of cottonseed oil and create a more visually appealing oil suitable for the retail bottled oil market. As noted above significant reduction of palmitic acid was observed in the HO-CSO and HS-CSO as a result of down-regulating the expression of either *ghFAD2-1* or *ghSAD-1*, but further reductions are desirable. Genetic down-regulation of *FatB* has resulted in significant reduction of palmitic acid in a number of oilseed crops including soybean (Kinney, Knowlton, Cahoon, and Hitz 1998) and canola (Pandian, Liu, Hurlestone, Singh, Salisbury, and Green 2004).

In contrast there are other applications where higher levels of palmitic and other saturated fatty acids are desired, such as in the production of confectionaries, cosmetics and lubricants. Increases in these saturates have been sought in other crops through introduction of genes encoding acyl-ACP thioesterase that specifically remove saturated acyl chains from the fatty acid synthesis pathway and make them available for inclusion in triglycerides without undergoing desaturation (Voelker et al. 1992). For example, transgenic expression in rapeseed of an acyl-ACP thioesterase gene from California bay (*Umbellularia californica*), a plant rich in the C12:0 saturate, lauric, acid in its oil, resulted in an accumulation of up to 50% lauric acid (Voelker et al. 1996). This led to the development of high-laurate canola, which has been used in the manufacturing

of soaps and detergents, and in food production, such as chocolate and some other confectionary products. Likewise, palmitic acid was substantially increased in model plants by expressing *FatB* acyl-ACP thioesterase genes derived from *Cuphea*, elm and nutmeg (Jones, Davies, and Voelker 1995; Voelker, Jones, Cranmer, Davies, and Knutzon 1997). An alternative approach that has been suggested for increasing palmitic acid involves down-regulation of β -keto acyl synthase II (KASII) which catalyses the conversion of palmitoyl-ACP to stearoyl-ACP in the plastids (Broun, Gettner, and Somerville 1999). However it was recently reported that in a T-DNA knockout mutant population of KASII in *Arabidopsis*, the inability to recover the homozygote mutant genotype suggested that KASII activity is essential for normal viability (Nguyen, Pidkowich, Whittle, and Shanklin 2006). Consistent with this, there were about 1/4 aborted ovules in siliques of T₁ plants. Strong seed-specific hairpin RNAi reductions in KASII also produced 1/4 aborted embryos. These results indicate that specific molecular genetic approaches to raising palmitic acid in cottonseed oil may have different levels of success.

3.5.2 Reduction of Cyclopropenoid Fatty Acids

Cottonseed oil contains small amounts of cyclopropenoid fatty acids (CPFAs) such as sterculic and malvalic acids. CPFAs are present, along with other fatty acids, in triacylglycerols and have been reported to range from about 0.6 to 1.5% of the total fatty acid content (Pandey and Suri 1982). These CPFAs elevate the melting point of fats in animals fed with whole cottonseed and cotton seed meal, by altering the ratio between stearic and oleic acids. As a result, CPFAs cause the hardening of fats in egg yolk and milk (Johnson, Pearson, Shenstone, and Fogarty 1967; Roehm, Lee, Wales, Polityka, and Sinnhuber 1970). The mechanism of action appears to be inhibition of desaturation of saturated fatty acids (Raju and Reiser 1972; 1973). In chicken, egg yolk discoloration and reduced hatchability are two detrimental effects, and consequently, industry limits the use of cottonseed meal and cottonseed oil in poultry diets (Phelps, Shenstone, Kemmerer, and Evans 1965). Sterculic and malvalic acids are derived from oleic acid through an intermediate, dihydrosterculic acid. As shown in Fig. 1, the enzyme responsible for converting oleic acid to dihydrosterculic acid is a methyltransferase, known as cyclopropane fatty acid synthase (CPA-FAS), which uses *S*-adenosyl-methionine as the methyl donor (Bao, Katz, Pollard, and Ohlrogge 2002; Bao, Thelen, Bonaventure, and Ohlrogge 2003), and a gene encoding this enzyme has been isolated from cotton (Liu et al. 2004).

3.5.3 Transgenic Production of Unusual Fatty Acids

In addition to the common edible fatty acids such as those found in cottonseed oil, members of the plant kingdom produce more than 200 unusual fatty acids, some of which comprise in excess of 70% of the total fatty acids in seed oil

(Badami and Patil 1981). Many of these unusual fatty acids have potential industrial applications as alternatives to petroleum-derived industrial feedstocks. While some of these plants containing high levels of unusual fatty acids are being domesticated, most others have limited agronomic potential. Therefore, because of the relatively low cost and renewable nature of oilseed crop production, cottonseed could be modified to produce unusual fatty acids.

In recent years, impressive progress has been made in isolating the key genes controlling the production of unusual fatty acids and their transformation into mainstream oilseed crops (Cahoon and Kinney 2005). Transgenic production of vernolic acid in cottonseed oil is one such example in exploring the opportunities in turning cottonseed oil into a highly valued industrial fatty acid source. Vernolic acid is an epoxy fatty acid which could be used as raw material for the production of resins, glues, plastics, polymers and other surface coatings. Currently epoxy fatty acids are produced through chemical epoxygenation of highly unsaturated vegetable oil or by synthesis from petrochemicals. It was discovered that the vernolic acid which comprises 70% of total fatty acids in seed oil of *Crepis palaestina* is catalysed by epoxygenation of linoleic acid by a divergent FAD2 desaturase-like enzyme (Lee, Lenman, Banas, Bafor, Singh, Schweizer, Nilsson, Liljenberg, Dahlqvist, Gummesson, Sjodahl, Green, and Stymne 1998). Expression of the cloned *C. palaestina* epoxygenase driven by a seed-specific napin promoter in cotton, led to accumulation of vernolic acid in cottonseed oil at a relatively low level of about 5%. Significant improvement was made by co-expressing this epoxygenase with a microsomal $\Delta 12$ fatty acid desaturase from *C. palaestina*, resulting in 2–3 fold increases in vernolic acid accumulation, with the highest level of 17% in cottonseeds (Zhou et al. 2006). This improvement is believed to be made by increase in epoxygenase substrate as the result of overexpression of the *C. palaestina* microsomal $\Delta 12$ fatty acid desaturase. However, this level of vernolic acid in cottonseed oil is still considerably less than the 70% found in *C. palaestina* where the epoxygenase gene originates from. Such a phenomenon is common to almost every other case where an unusual fatty acid has been made by a transgenic approach (Suh, Schultz, and Ohlrogge 2002; Singh, Zhou, Liu, Stymne, and Green 2005). Hence, the current challenge is to obtain high levels of unusual fatty acid accumulation in cottonseed leading to an economically viable outcome. Further, while the plant physiology of the epoxy cotton is still being studied, the accumulation of unusual fatty acids in oilseed crops is often associated with impaired seed viability, poor germination, and seriously compromised plant growth (Cahoon and Kinney 2005; Murphy 2006). It has been postulated that such problems may be caused by the accumulation of these unusual fatty acids in the membrane phospholipids in transgenic plants, therefore contributing to the failure to accumulate significant amounts of the new fatty acids, as well as the problems with seed viability (Millar, Smith, and Kunst 2000; Thomaus, Carlsson, and Stymne 2001). Plants that normally accumulate these unusual fatty acids have evolved an ability to exclude the unusual fatty acids from cell membranes. For example, *C. palaestina* which normally accumulates a high

level of vernolic acid in its seed oil is likely to possess additional acyltransferases which have substrate-specificity towards vernolic acid and are capable of channeling of vernolic acid into TAG and sequestering it away from membrane lipids (Singh et al. 2005; Dyer and Mullen 2005). It is generally believed that diacylglycerol acyltransferase (DGAT), the last enzyme in the Kennedy pathway has a strong influence on both the quantitative and qualitative accumulation of fatty acids in TAG biosynthesis. For example, a DGAT from cocoa has strong preference for utilizing stearyl-CoA as the acyl donor for TAG synthesis (Griffiths and Harwood 1991) and some DGATs from erucic acid-producing rapeseed show similar preference for erucoyl-CoA (Bernert and Frentzen 1990). As well as this qualitative selectivity, DGAT also seems to play a quantitative role in TAG synthesis, since overexpression of DGAT in a number of plants including tobacco, *Arabidopsis*, soybean and canola has resulted in enhanced oil content (Jako, Kumar, Wei, Zou, Barton, Giblin, Covello, and Taylor 2001; Bourier-Nave, Benveniste, Oelkers, Sturley, and Schaller 2000; Bennett, Lardizabel, Keithly, Mai, Hill, Ream, Wagner, Colletti, and Crow 2004; Weselake, Shah, Taylor, Harwood, Laroche, Moloney, Rakow, and Raney 2006).

3.6 Reduction of Gossypol in Cottonseed

Cotton contains a number of terpenoid phytoalexins of which gossypol is most notable (Cherry et al. 1981; Cherry 1983). These phytoalexins accumulate in sub-epidermal glands in the aerial parts of the plant and the seed, and play a critical role in their resistance to potential pathogens. Gossypol is particularly toxic to non-ruminant animals and has inhibitory effects on male fertility when cottonseed is used for feeding either directly or as a meal following oil extraction. Gossypol occurs in either free or bound form, and the former is toxic. Sudweeks (2002) suggested that 18 mg of free gossypol (0.1%) is the maximum that should be fed to dairy cows. During the oil extraction process, some gossypol is deactivated through moist heating which causes the formation of a double bond between the ϵ -amino group of lysine and the aldehyde group in the gossypol. Although effective in reducing the toxicity of gossypol, the binding of gossypol reduces the amount of soluble protein and bioactive lysine in the meal (Zarins and Cherry 1981). Some recent advances in oil processing methods allow gossypol to be effectively removed, but with significantly increased cost and a reduction of oil yield.

Early attempts to remove gossypol focused on development of mutant cotton plants that lacked the gossypol secreting glands (McMichael 1954; 1959). However glandless cotton plants proved to be highly susceptible to insects and fungal diseases during the vegetative phase, and so subsequent strategies have focused on molecular methods targeting seed-specific reductions in gossypol. In recent years, δ -cadinene synthase catalyzing the first

committed step in cadinene-type sesquiterpene biosynthesis was identified as the rate-limiting factor for gossypol biosynthesis (Benedict, Alchanati, Harvey, Liu, Stipanovic, and Bell 1995). The subsequent isolation of genes encoding this enzyme (Meng, Jia, Liu, Liang, Heinsteins, and Chen 1999) paved the way for eventual elimination of gossypol in a seed-specific manner. Despite initial disappointment with transforming cotton with an antisense- δ -cadinene synthase (Townsend, Poole, Blake, and Llewellyn 2005), subsequent RNAi silencing of this gene led to a drastic reduction of gossypol levels in a seed-specific manner, without reducing this compound and other related terpenoids in somatic tissues (Sunikumar et al. 2006). Removing gossypol will permit use of cottonseed protein in feed rations for non-ruminant animals, such as poultry, swine and fish. For the traditional market of the beef industry, eliminating gossypol should lead to the increased availability of lysine in the meal. Reduction or elimination of gossypol is highly desirable in cottonseeds having either traditional or genetically modified fatty acid compositions.

3.7 Improvement in Minor Constituents

3.7.1 Phytosterols

Cottonseed oil is a good source of phytosterols, which have been shown to lower blood cholesterol levels by 10–15% and may therefore help reduce the risk of cardiovascular disease (Ling and Jones 1995). In addition, plant sterols may possess anti-cancer, anti-atherosclerosis, anti-inflammation and anti-oxidation activities (Berger, Jones, and Abumweis 2004). The unsaponifiable matter in cottonseed oil ranges from 0.5–0.7% in total and is dominated by phytosterols, mainly β -sitosterol and campesterol (Cherry et al. 1981; Cherry and Leffler 1984; Kritchevsky 1998). Vegetable oils containing substantially higher levels of phytosterols/stanols have been generated by expressing the gene encoding the rate-limiting enzyme for sterol synthesis, hydroxymethylglutaryl-CoA reductase, in seeds of crop plants (Venkatramesh, Karunanandaa, Gunter, Thorne, and Crowley 2000; Harker, Hellyer, Clayton, Duvoix, Lanot, and Safford 2003). Development of cottonseed oil with increased phytosterols might provide added nutritional value suitable for particular end-product applications.

3.7.2 Vitamin E

The potential exists to produce elevated levels of tocopherols (Vitamin E) in cottonseed oil, which are the most important lipid soluble antioxidants in the cell membranes acting through a free radical scavenging mechanism. Tocopherols have been implicated in decreased risk of cardiovascular disease and some forms of cancer, improved immune functions, and in slowing the progress of some degenerative diseases in humans (Fryer 1992; Traber and Sies 1996). They

are also strong antioxidants which increase oxidative stability of cottonseed oil. The level of natural tocopherols in refined cottonseed oil is 700–800 ppm, mainly α - and γ -tocopherols, of which α -tocopherol possesses the highest Vitamin E activity (Smith and Creelman 2001). Shintani and DellaPenna (1998) have reported the cloning of the final enzyme in plant α -tocopherol synthesis, γ -tocopherol methyltransferase (γ -TMT). Overexpression of γ -TMT in *Arabidopsis* seeds shifted tocopherol synthesis in favor of α -tocopherol. Seeds of the lines overexpressing the largest amounts of γ -TMT contained greater than 95% of their total tocopherol pool as α -tocopherol. This indicated that the enhancement of α -tocopherol levels in oilseed crops such as cotton could be possible through genetic engineering.

3.7.3 Neuroactive N-Acylethanolamines

N-Acylethanolamines (NAEs) are bioactive fatty acid derivatives found in small quantities in all plants and animals (Chapman 2004). In animal systems these compounds regulate a wide array of physiological processes including feeding behavior, neurotransmission, pain perception, immune function and embryo implantation (Felder, Dickason-Chesterfield, and Moore 2006). These lipids act to regulate mammalian physiology in part through receptors like the cannabinoid receptors and have become known as endocannabinoid signaling compounds (Ligresti, Cascio, and Di Marzo 2005). NAEs are present in ppm quantities in desiccated oilseeds (Venables, Waggoner, and Chapman 2005), and substantial quantities of these neuroactive NAEs can be recovered in cottonseed processing streams (Chapman, Venables, Dian, and Gross 2003). As more is learned about the therapeutic potential of different NAE types, cottonseed may represent a natural source of these bioactive lipids and this may become a valuable nutraceutical co-product of cottonseed oil, somewhat similar to tocopherols. In the context of fatty acid modification, it should be pointed out that NAE activities in animals vary based on the length and degree of unsaturation, so modification of fatty acid composition of the cottonseed oil may facilitate the development of novel therapeutic NAE compositions as well.

4 Summary and Perspectives

Conventional plant breeding coupled with biotechnology can be exploited to produce cottonseed oil with improved nutritional values and broad industrial applications. Conventional plant breeding involves selecting amongst variability in existing germplasm or randomly induced mutations. Modern biotechnology is a powerful tool which can extend the ability of geneticists and plant breeders to modify functional properties of oilseed crops, guided by detailed understanding of lipid biochemistry and modern molecular biology methodologies.

Significant progress has been made in the genetic modification of cottonseed oil for improved nutritional value and functional properties. High-oleic

cottonseed oil has good stability and could be directly used in the food service sector without hydrogenation. Substantial reduction of palmitic acid and cyclopropenoid fatty acids in the high-oleic oil can further enhance cottonseed oil's nutritional elements and competitiveness against other oilseeds, such as soybean and canola. Both high stearic and cocoa butter substitute (combination of high-oleic and high-stearic) types of cottonseed oil have great potential for use as solid fat applications, especially in the highly valued confectionary industry. Beyond the modification of fatty acid chain length and saturation level, RNAi technology has also been successfully used in reducing nutritionally undesirable compounds such as gossypol. In addition, with rapid progress in the identification and isolation of enzymes for the production of unusual fatty acids, accelerated by functional genomics, cottonseed oil could also be modified to produce fatty acids with valuable industrial applications, such as epoxy fatty acids. A number of oil-soluble bioactive compounds, such as phytosterols, tocopherols and neuroactive *N*-Acylethanolamines might also be enhanced or produced through genetic engineering. New technologies are emerging in the area of manipulating TAG structure and oil yield. Ultimately, it will be possible to produce cottonseed oils tailored to users' specific needs through genetic engineering. As cottonseed is a by-product of the more valuable cotton fiber, the improvement of nutritional value and functional properties of a cottonseed oil could add further value while maintaining the superior fiber quality and yield and agronomic performance.

Recently, avenues to increase utilization of renewable energy resources have attracted significant attention. Cottonseed oil represents one of the most widely available feedstocks for biodiesel and oleochemical industry. Currently a large proportion of cottonseeds are used as animal feed. Because of the new and lively interest in biodiesel this situation may change. Demand for use in industrial purposes has risen more quickly than that for food or feed. The high-oleic and high-stearic cottonseed oil initially developed for the health food industry might find their ways into biodiesel use because of their enhanced cetane value and cold flow properties. However, the production of novel fatty acids in plant seed oil, including cottonseed oil, has so far still been at the proof-of-concept stage. Despite initial optimism, transgenic production of an industrial fatty acid by expressing a single epoxygenase has turned out to be more complex than was first thought. It appears that our understanding of even the basic pathway of TAG biosynthesis is far from complete and that there are probably multiple rate-limiting factors rather than just one. Current work has mostly focused on the production of novel fatty acids by introducing the catalytic enzymes responsible for their synthesis. However, it is likely that there is a complex interplay among the introduced fatty acid and the endogenous enzymes involved in the TAG assembly, such as acyltransferases. These systems need to be working in concert with an efficient channeling of the novel fatty acids into the TAG, for the successful production of high levels of unusual fatty acids in transgenic cottonseed oil (Cahoon and Kinney 2005; Dyer and Mullen 2005).

It is expected that the current hurdles with transgenic production of high value novel fatty acids could be overcome through a better understanding of the biochemical pathways and the exploitation of the rapid development in areas of functional genomics and molecular biology methodology. A comprehensive catalogue of *Arabidopsis* lipid genes has been established (Beisson, Koo, Ruuska, Schwender, Pollard, Thelen, Paddock, Salas, Savage, Milcamps, Mhaske, Cho, and Ohlrogge 2003; <http://www.plantbiology.msu.edu/lipids/genesurvey/index.html>) and could be used to identify cotton orthologues. The development of microarrays enables the global analysis of gene expression in developing seeds and leads to a better understanding of the complex metabolic networks that control fatty acid biosynthesis TAG assembly during the oil accumulation process (Ruuska, Girke, Benning, and Ohlrogge 2002; Wu, Machdo, White, Llewellyn, and Dennis 2006). Through such technology, the simultaneous measurement of thousands of transcripts, and RNA populations in developing embryos from different sources can be directly compared. Reverse genetics tools including transposon or T-DNA tagging and expression of RNAi transgenes that downregulate the expression of endogenous genes can also be used in studying mechanisms of oil biosynthesis.

Facilitated by a good understanding of the biochemical pathway and the availability of DNA sequence data, certain traits can also be produced by a non-transgenic route. This is a particularly welcome development, given sometimes negative consumer sentiment about GM crops. A newly developed technology, targeting induced local lesions in genomes (TILLING) (McCallum, Comai, Greene, and Henikoff 2000), combines chemical mutagenesis with mutation screening using gene-specific PCR primers, allowing the isolation of missense and nonsense mutant alleles of targeted genes. TILLING can be automated in a high throughput system and mutants with lesions in a specific gene can be identified. Due to the tetraploid nature of cotton, at least two copies of any given candidate gene can be expected. Detailed sequence data in polymorphic regions of a gene could be used to identify the orthologous gene loci and guide the design of specific PCR primers to detect mutations with different subgenomic origin. Combination of two such mutants could lead to cotton plants showing the desirable fatty acid composition as previously seen in the RNAi transgenic plants.

References

- Andersen, P.C., Hill, K., Gorbet, D.W., Brodbeck, B.V. (1998) Fatty acid and amino acid profiles of selected peanut cultivars and breeding lines. *J. Food Composition Anal.* 11, 100–111.
- Athenstaedt, K. and Daum, G. (2006) The life cycle of neutral lipids: Synthesis, storage and degradation. *Cell. Mol. Life Sci.* 63, 1355–1369.
- Auld, D., Bechere, E., Davis, J., Seip, L. and Brown, J. (2006) Lint, cottonseed oil, or biodiesel? -Breeding cotton for the next decade. The ASA-CSSA-SSSA Annual International Meetings, Indianapolis, Nov. 12–16, 2006.

- Badami, R.C. and Patil, K.B. (1981) Structure and occurrence of unusual fatty acids in minor seed oils. *Lipid Res.* 19, 119–153.
- Bao, X., Katz, S., Pollard, M. and Ohlrogge, J. (2002) Carboxylic fatty acids in plants: biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of *Sterculia foetida*. *Proc. Natl. Acad. Sci. USA* 99, 7172–7177.
- Bao, X., Thelen, J.J., Bonaventure, G. and Ohlrogge, J.B. (2003) Characterization of cyclopropane fatty-acid synthase from *Sterculia foetida*. *J. Biol. Chem.* 278, 12846–12853.
- Beisson, F., Koo, A.J.K., Ruuska, S., Schwender, J., Pollard, M., Thelen, J.J., Paddock, T., Salas, J.J., Savage, L., Milcamps, A., Mhaske, V.B., Cho, Y.H. and Ohlrogge, J.B. (2003) *Arabidopsis* genes involved in acyl lipid metabolism -A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.* 132, 681–697.
- Benedict, C.R., Alchanati, I., Harvey, P.J., Liu, J., Stipanovic, R.D. and Bell, A.A. (1995) The enzymatic formation of δ -cadinene from farnesyl diphosphate in extracts of cotton. *Phytochem.* 39, 327–331.
- Bennett, K.A., Lardizabel, K.D., Keithly, G.E., Mai, J.T., Hill, E.K., Ream, J.E., Wagner, N.N., Colletti, R.F. and Crow, L.J. (2004) Modification of seed oil content in soybean (*Glycine max*) by expression of a *Mortierella ramanniana* diacylglycerol acyltransferase. *Lake Buena Vista, FL*, July 24–28, 2004.
- Berger, A., Jones, P.J.H. and Abumweis, S.S. (2004) Plant sterols: factors affecting their efficacy and safety as functional food ingredients. *Lipids in Health and Disease* 3, 5.
- Bernerth, R. and Frentzen, M. (1990) Utilization of erucyl-CoA by acyltransferases from developing seeds of *Brassica napus* (L.) involved in triacylglycerol biosynthesis. *Plant Sci.* 67, 21–28.
- Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366.
- Bonanome, A. and Grundy, S.M. (1988) Effect of dietary stearic acid on plasma cholesterol and lipoprotein levels. *N. Engl. J. Med.* 318, 1244–1248.
- Bouvier-Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and Schaller, H. (2000) Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase. *Eur. J. Biochem.* 267, 85–96.
- Broun, P., Gettner, S. and Somerville, C. (1999) Genetic engineering of plant lipids. *Annu. Rev. Nutr.* 19, 197–216.
- Browse, J. and Ohlrogge, J. (1995) Lipid biosynthesis. *Plant Cell* 7, 957–970.
- Buhr, T., Sato, S., Ebrahim, F., Xing, A.Q., Zhou, Y., Mathiesen, M., Schweiger, B., Kinney, A., Staswick, P. and Clemente, T. (2002) Nuclear localization of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J.* 30, 155–163.
- Burton, J.W., Miller, J.F., Vick, B.A., Scarth, R. and Holbrook, C.C. (2004) Altering fatty acid composition in oil seed crops. *Adv. Agron.* 4, 273–306.
- Cahoon, E.B. (2003) Genetic enhancement of soybean oil for industrial uses: prospects and challenges. *AgBioForum* 6, 11–13.
- Cahoon, E.B. and Kinney, A.J. (2005) Production of vegetable oils with novel properties: using genomic tools to probe and manipulate fatty acid metabolism. *Eur. J. Lipid Sci. Technol.* 107, 239–243.
- Chapman, K.D. (2004) Occurrence, metabolism, and prospective functions of N-acyl ethanolamines in plants. *Prog. Lipid Res.* 43, 302–327.
- Chapman, K.D., Austin-Brown, S., Sparace, S.A., Kinney, A.J., Ripp, K.G., Pirtle, I.L. and Pirtle, R.M. (2001) Transgenic cotton plants with increased seed oleic acid content. *J. Am. Oil Chem. Soc.* 78, 941–947.
- Chapman, K.D., Neogi, P.B., Hake, K.D., Stawska, A.A., Speed, T.R., Cotter, M.Q., Garrett, D.C., Kerby, T., Richardson, C.D., Ayre, B.G., Ghosh, S. and Kinney, A.J. (2008) Reduced oil accumulation in cottonseeds transformed with a *Brassica* nonfunctional allele of a delta-12 fatty acid desaturase (*FAD2*). *Crop Sci* 48, 1470–1481.

- Chapman, K.D., Venables, B.J., Dian, E.E. and Gross, G.W. (2003) Identification and quantification of neuroactive *N*-acylethanolamines in cottonseed processing fractions. *J. Am. Oil Chem. Soc.* 80, 223–229.
- Cherry, J.P. (1983) Cottonseed oil. *J. Am. Oil Chem. Soc.* 60, 360–367.
- Cherry, J.P. and Leffler, H.R. (1984) Chapter 13: Seed. In: R.J. Kohel, and C.F. Lewis (Eds.) *Cotton* No. 24 in Agronomy series; American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America. Madison, WI, pp. 511–569.
- Cherry, J.P., Kohel, R.J., Jones, L.A. and Powell, W.H. (1981) Cottonseed quality: factors affecting feed and food uses. In: J.M. Brown (Ed.) *Proceedings Beltwide Cotton Production Research Conference*. Memphis, TN, pp. 266–283.
- Cherry, J.P., Kohel, R.J., Jones, L.A. and Powell, W.H. (1986) Food and feed quality of cottonseed. In: J.R. Mauney and J.McD. Stewart (Eds.) *Cotton Physiology*. The Cotton Foundation: Memphis, TN, pp. 557–595.
- Coppock, C.E., Lanham, J.K. and Horner, J.L. (1987) A review of the nutritive value and utilization of whole cottonseed, cottonseed meal and associated by-products by dairy cattle. *Anim. Feed Sci. Technol.* 18, 89–129.
- Crandall, L. (2002) Bioplastics: A burgeoning industry. *INFORM* 13, 626–630.
- Downey, R.K. and Craig, B.M. (1964) Genetic control of fatty acid biosynthesis in rapeseed (*Brassica napus* L.). *J. Am. Oil Chem. Soc.* 41, 475–478.
- Dyer, J.M. and Mullen, R.T. (2005) Development and potential of genetically engineered oilseeds. *Seed Sci. Res.* 15, 255–267.
- Felder, C.C., Dickason-Chesterfield, A.K. and Moore, S.A. (2006) Cannabinoids biology: the search for new therapeutic targets. *Mol. Interv.* 6, 149–161.
- Fernández-Martínez, J. del Río, M., de Hare, A. (1993) Survey of safflower (*Carthamus tinctorius* L.) germplasm for variants in fatty acid composition and other seed characters. *Euphytica* 69, 115–122.
- Fofana, B., Cloutier, S., Duguid, S., Ching, J. and Rampitsch, C. (2006) Gene expression of stearoyl-ACP desaturase and $\Delta 12$ fatty acid desaturase 2 is modulated during seed development of flax (*Linum usitatissimum*). *Lipids* 41, 705–712.
- Fofana, B., Duguid, S. and Cloutier, S. (2004) Cloning of fatty acid biosynthetic genes beta-ketoacyl CoA synthase, fatty acid elongase, stearoyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum* L.) seeds. *Plant Sci.* 166, 1487–1496.
- Fourmann, M., Barrel, P., Renard, M., Pelletier, G., Delourme, R. and Brunel, D. (1998) The two genes homologous to *Arabidopsis* FAE1 co-segregate with the two loci governing erucic fatty acid content in *Brassica napus*. *Theor. Appl. Genet.* 96, 852–858.
- Fryer, M.J. (1992) The antioxidant effects of thylakoid vitamin E (α -tocopherol). *Plant Cell Environ.* 1, 211–215.
- Girke, Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124, 1570–1581.
- Gotmare, V., Singh, P., Mayee, C.D., Deshpande, V. and Bhagat, C. (2004) Genetic variability for seed oil content and seed index in some wild species and perennial races of cotton. *Plant Breed.* 123, 207–208.
- Graef, G.L., Fehr, W.R. and Hamond, E.G. (1985) Inheritance of three stearic acid mutants of soybean. *Crop Sci.* 25, 1076–1079.
- Green, A.G. and Dribnenki, J.C.P. (1994) Linola -A new premium polyunsaturated oil. *Lipid Technol.* 6, 29–33.
- Green, A.G. and Marshall, D.R. (1984) Isolation of induced mutants in linseed (*Linum usitatissimum*) having reduced linolenic acid content. *Euphytica* 33, 321–328.
- Griffiths, G. and Harwood, J.L. (1991) The regulation of triacylglycerol biosynthesis in cocoa (*Theobroma cacao* L.) *Planta* 184, 279–284.

- Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296.
- Harker, M., Hellyer, A., Clayton, J.C., Duvoix, A., Lanot, A. and Safford, R. (2003) Co-ordinate regulation of sterol biosynthesis enzyme activity during accumulation of sterols in developing rape and tobacco seed. *Planta* 216, 707–715.
- Hawkins, D.J. and Kridl, J.C. (1998) Characterization of acyl-ACP thioesterases of mangosteen (*Garcinia mangostana*) seed and high levels of stearate production in transgenic canola. *Plant J.* 13, 743–752.
- Hill, K. (2000) Fats and oils as oleochemical raw materials. *Pure Appl. Chem.* 72, 1255–1264.
- Howell, R.W. (1971) Breeding for improved oilseeds. *J. Am. Oil Chem. Soc.* 48, 492–494.
- Jako, C., Kumar, A., Wei, Y., Zou, J., Barton, D.L., Giblin, E.M., Covello, P.S. and Taylor, D.C. (2001) Seed-specific over-expression of an *Arabidopsis* encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.* 126, 861–874.
- Johnson, A.R., Pearson, J.A., Shenstone, F.S. and Fogarty, A.C. (1967) Inhibition of stearic to oleic acid by cyclopropene fatty acids. *Nature* 214, 1244–1245.
- Jones, A., Davies, H.M. and Voelker, T.A. (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell* 7, 359–371.
- Jones, L. and King, C. (1993) *Cottonseed Oil*. National Cottonseed Products Associations, Inc. and the Cotton Foundation, Memphis, TN, USA.
- Kinney, A.J. (1996a) Improving soybean seed quality. *Nature Biotechnol.* 14, 946.
- Kinney, A.J. (1996b) β -ketocyl-ACP synthetase II genes from plants. US patent 5,500,361.
- Kinney, A.J. and Clemente, T.E. (2005) Modifying soybean oil for enhanced performance in biodiesel blends. *Fuel Proc. Techn.* 86, 1137–1147.
- Kinney, A.J., Knowlton, S., Cahoon, E.B. and Hitz, W.D. (1998) Re-engineering oilseed crops to produce industrially useful fatty acids. In: J. Sánchez, E. Cerdá-Olmedo and E. Martínez-Force (Eds) *Advances in Plant Lipid Research*. University of Seville Press, Seville, pp. 623–628.
- Kleingartner, L.W. (2002) NuSun sunflower oil: Redirection of an industry. In: J. Janick and A. Whipkey (Eds) *Trends in New Crops and New Uses*. ASHS Press, Alexandria, VA. pp 135–138.
- Knutzon, D.S., Thompson, G.A., Radke, S.E., Johnson, W.B., Knauf, V.C. and Kridl, J.C. (1992) Modification of *Brassica* seed oil by antisense expression of a stearyl-acyl carrier protein desaturase gene. *Proc. Natl. Acad. Sci. USA* 89, 2624–2628.
- Kris-Etherton, P.M., Derr, J., Mitchell, D.C., Mustard, V.A., Russell, M.E., McDonnell, E.T. Salabsky, D. and Pearson, T.A. (1993a) The role of fatty acid saturation on plasma lipids, lipoproteins and apolipoproteins: I. Effects of whole food diets high in cocoa butter, olive oil, soybean oil, dairy butter and milk chocolate on plasma lipids of young men. *Metabolism* 42, 121–129.
- Kris-Etherton, P.M., Mustad, V. and Derr, J.A. (1993b) Effects of dietary stearic acid on plasma lipids and thrombosis. *Nutrition-today (USA)*. 28, 30–38.
- Kristott, J. (2003) High-oleic oils: how good are they for frying? *Lipid Techn.* 15, 29–32.
- Kritchevsky, D. (1998) Phytosterols. In: D. Kritchevsky and C. Bonfield (Eds) *Dietary Fiber in Health and Disease*. Plenum Press, New York, pp 235–243.
- Lee, M., Lenman, M., Banas, A., Bafar, M., Singh, S., Schweizer, M., Nilsson, R., Liljenberg, C., Dahlqvist, A., Gummesson, P.O., Sjodahl, S., Green, A. and Stymne, S. (1998) Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science* 280, 915–918.
- Lichtenstein, A.H., Appel, L.J., Brands, M., Carnethon, M., Daniels, S., Franch, H.A., Franklin, B., Kris-Etherton, P., Harris, W.S., Howard, B., Karanja, N., Lefevre, M., Rudel, L., Sacks, F., Van Horn, L., Winston, M. and Wylie-Rosett, J. (2006) Diet and lifestyle recommendations revision 2006: A scientific statement from the American heart association nutrition committee. *Circulation* 114, 82–96.

- Ligresti, A., Cascio, M.G. and Di Marzo, V. (2005) Endocannabinoid metabolic pathways and enzymes. *Curr. Drug Targets CNS Neurol. Disord.* 4, 615–623.
- Lindsey, S., Benattar, J., Pronczuk, A. and Hayes, K.C. (1990) Dietary palmitic acid (16:0) enhances high density lipoprotein cholesterol and low density lipoprotein receptor mRNA abundance in hamsters. *Exp. Biol. Med.* 195, 261–169.
- Ling, W.H. and Jones, P.J.H. (1995) Dietary phytosterols: a review of metabolism, benefits and side effects. *Life Sci.* 57, 195–206.
- List, G.R., Steidley, K.R. and Neff, W.E. (2000) Commercial spreads formulation, structure and properties. *INFORM* 11, 980–986.
- Liu, Q., Bao, X., Singh, S., Pons, G., Hurlstone, C., Dowd, C., McFadden, H., Pollard, M., Ohlrogge, J.B. and Green, A. (2004) Studies on cotton cyclopropane fatty acid synthase. In: *16th International Symposium on Plant Lipids*. Budapest, Hungary, June 1–6 2004.
- Liu, Q., Singh, S., Sharp, P.J., Green, A.G. and Marshall, D.R. (1996) Nucleotide sequence of a cDNA from *Gossypium hirsutum* encoding a stearyl-acyl carrier protein desaturase (Accession No. X95988) (OGR96-012). *Plant Physiol.* 110, 1435.
- Liu, Q., Singh, S.P. and Green, A.G. (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiol.* 129, 1732–1743.
- Liu, Q., Singh, S.P., Brubaker, C.L. and Green, A.G. (1999a) Cloning and sequence analysis of a novel member (accession No. Y10112) of the microsomal ω -6 fatty acid desaturase family from cotton (*Gossypium hirsutum*). *Plant Physiol.* 120, 339.
- Liu, Q., Singh, S.P., Brubaker, C.L., Sharp, P.J., Green, A.G. and Marshall, D.R. (1999b) Molecular cloning and expression of a cDNA encoding a microsomal ω -6 fatty acid desaturase in cotton (*Gossypium hirsutum* L.). *Aust. J. Plant Physiol.* 26, 101–106.
- Lukonge, E., Labuschagne, M.T. and Hugo, A. (2007) The evaluation of oil and fatty acid composition in seed of cotton accessions from various countries. *J. Sci. Food Agri.* 87, 340–347.
- McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123, 439–442.
- McMichael, S.C. (1954) Glandless boll in upland cotton and its use in the study of natural crossing. *Agron. J.* 46, 527–528.
- McMichael, S.C. (1959) Hope cotton, a source of cottonseed free of gossypol pigments. *Agron. J.* 51, 630.
- Meng, Y.-L., Jia, J.-W., Liu, C.-J., Liang, W.-Q., Heinsteins, P. and Chen, X.-Y. (1999) Coordinated accumulation of (+)- δ -cadinene synthase mRNAs and gossypol in developing seeds of *Gossypium hirsutum* and a new member of the cad1 family from *G. arboreum*. *J. Nat. Prod.* 62, 248–252.
- Millar, A., Smith, M. and Kunst, L. (2000) All fatty acids are not equal: discrimination in plant membrane lipids. *Trends Plant Sci.* 5, 95–101.
- Mohamed, O.E., Satter, L.D., Grummer, R.R. and Ehle, F.R. (1988) Influence of dietary cottonseed and soybean on milk production and composition. *J. Dairy Sci.* 71, 2677.
- Möllers, C. (2002) Development of oleic acid oilseed rape. In: *Presentation at the 8th International Conference for Renewable Resources and Plant Biotechnology*. Biotechnology NAROSSA 2002. Magdeburg, June 10–11, 2002.
- Mozaffarian, D., Katan, M.B., Ascherio, A., Stampfer, M.J. and Willett, W.C. (2006) *Trans* fatty acids and cardiovascular disease. *N. Engl. J. Med.* 354, 1601–1613.
- Murphy, D.J. (2006) Molecular breeding strategies for the modification of lipid composition. *In Vitro Cell. Dev. Biol. -Plant* 42, 89–99.
- Nguyen, H.T., Pidkowich, M.S., Whittle, E. and Shanklin, J. (2006) *FAB1*, encoding β -ketoacyl-ACP synthase II, is an essential gene in *Arabidopsis*. *17th International Symposium on Plant Lipids*. East Lansing, MI, July 16–21, 2006.
- Nicolosi, R.J. and Rogers, E.J. (1997) Regulation of plasma lipoprotein levels by dietary triglycerides enriched with different fatty acids. *Med. Sci. Sports Exerc.* 29, 1422–1428.

- O'Brien, R.D. (2002) Cottonseed Oil. In: F.D. Gunstone (Ed.) *Vegetable Oils in Food Technology: Composition, Properties and Uses*. Blackwell Publishing, Oxford, pp. 203–230.
- Ohlrogge, J.B. and Jaworski, J.G. (1997) Regulation of fatty acid synthesis. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 48, 109–136.
- Oomen, C.M., Ocke, M.C., Feskens, E.J., van Erp-Baart, M.A., Kok, F.J. and Kromhout, D. (2001) Association between *trans* fatty acid intake and 10-year risk of coronary heart disease in the Zutphen Elderly Study: A perspective population-based study. *Lancet* 357, 746–751.
- Osorio, J., Fernandez-Martínez, J., Mancha, M. and Garcés, R. (1995) Mutant sunflower with high concentration of saturated fatty acids in the oil. *Crop Sci.* 35, 739–742.
- Pandey, S.N. and Suri, L.K. (1982) Cyclopropenoid fatty acid content and iodine value of crude oils from Indian cottonseed. *J. Am. Oil Chem. Soc.* 59, 99–101.
- Pandian, A., Liu, Q., Hurlestone, C., Singh, S., Salisbury, P. and Green, A. (2004) Development of nutritionally superior *Brassica napus* and *B. juncea* oils using RNAi-mediated gene silencing. 4th International Crop Science Congress. Brisbane, Australia. Sept. 26–Oct. 01, 2004.
- Percival, A.E. and Kohel, R.J. (1990) Distribution, collection and evaluation of *Gossypium*. *Ad. Agron.* 44, 225–256.
- Phelps, R.A., Shenstone, F.S., Kemmerer, A.R. and Evans, R.J. (1965) A review of cyclopropenoid compounds: Biological effects of some derivatives. *Poult. Sci.* 44, 358–394.
- Pirtle, I.L., Kongcharoensuntorn, W., Nampaisansuk, M., Knesek, J.E., Chapman, K.D. and Pirtle, R.M. (2001) Molecular cloning and functional expression of the gene for a cotton Δ -12 fatty acid desaturase (FAD2). *Biochim. Biophys. Acta* 1522, 122–129.
- Rahman, S.M., Takagi, Y., Miyamoto, K., Kawakita, T. (1995) High stearic acid soybean mutant induced by X-ray irradiation. *Biosci. Biotechnol. Biochem.* 59, 922–923.
- Raju, P.K. and Reiser, R. (1972) Inhibition of fatty acyl desaturase by cyclopropene fatty acids. *J. Biol. Chem.* 247, 3700–3701.
- Raju, P.K. and Reiser, R. (1973) Hepatic stearoyl-CoA desaturase activity in mice as affected by early postnatal dietary cyclopropene fatty acids I. *J. Nutr.* 103, 904–907.
- Roehm, J.N., Lee, D.J., Wales, J.H., Polityka, S.D. and Sinnhuber, P.D. (1970) The effect of the dietary sterculic acid on the hepatic lipids of rainbow trout. *Lipids* 5, 80–84.
- Ruuska, S.A., Girke, T., Benning, C. and Ohlrogge, J.B. (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14, 1191–1206.
- Shenstone, F.S. and Vickery, J.R. (1961) Occurrence of cyclo-propene acids in some plants of the order malvales. *Nature* 190, 68–169.
- Shintani, D. and DellaPenna, D. (1998) Elevating the vitamin E content of plants through metabolic engineering. *Science* 228, 2098–2100.
- Singh, S., Zhou, X.-R., Liu, Q., Stymne, S. and Green, A.G. (2005) Metabolic engineering of new fatty acids in plants. *Cur. Opin. Plant Biol.* 8, 197–203.
- Slocombe, S.P., Pianelli, P., Fairbairn, D., Bowra, S., Hatzopoulos, P., Tsiantis, M. and Murphy, D.J. (1994) Temporal and tissue-specific regulation of a *Brassica napus* stearoyl-acyl carrier protein desaturase gene. *Plant Physiol.* 104, 1167–1176.
- Smith, C.W. and Creelman, R.A. (2001) Vitamin E concentration in upland cottonseeds. *Crop Sci.* 41, 577–579.
- Stoll, C., Lühs, W., Zarhloul, M.K. and Friedt, W. (2005) Genetic modification of saturated fatty acids in oilseed rape (*Brassica napus*). *Eur. J. Lipid Sci. Techn.* 107, 4244–4248.
- Sudweeks, E.M. (2002) Feeding whole cottonseed to dairy cows and replacements. Texas A&M University, College Station, TX, U.S.A. Publication 13277101 -cottonsd.wp6.
- Suh, C., Schultz, D.J. and Ohlrogge, J.B. (2002) What limits production of unusual monoenoic fatty acids in transgenic plants? *Planta* 215, 584–595.
- Sunilkumar, G., Campbell, L.M., Hossen, M., Connell, J.P., Hernandez, E., Reddy, A.S., Smith, C.W. and Rathore, K.S. (2005) A comprehensive study of the use of a homologous

- promoter in antisense cotton lines exhibiting a high seed oleic acid phenotype. *Plant Biotechnol. J.* 3, 319–330.
- Sunilkumar, G., Campbell, L.M., Puckhaber, L., Stipanovic, R.D. and Rathore, K.S. (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc. Natl. Acad. Sci. USA* 103, 18054–18059.
- Tarrago-Trani, M.T., Philips, K.M., Lemar, L.E. and Holden, J.M. (2006) New and existing oils and fats used in products with reduced *trans*-fatty acid content. *J. Am. Diet. Assoc.* 106, 867–880.
- Thomaeus, S., Carlsson, A.S. and Stymne, S. (2001) Distribution of fatty acids in polar and neutral lipids during seed development in *Arabidopsis thaliana* genetically engineered to produce acetylenic, epoxy and hydroxy fatty acids. *Plant Sci.* 161, 997–1003.
- Thompson, G.A. and Li, C. (1997) Altered fatty acid composition of membrane lipids in seeds and seedling tissues of high-saturate canola. In: J.P. Williams, M.U. Khan, and N.M. Lem (Eds.) *Physiology, Biochemistry and Molecular Biology of Plant Lipids*. Kluwer Academic Publishers, Dordrecht, pp. 313–315.
- Töpfer, R., Martini, N. and Schell, J. (1995) Modification of plant lipid synthesis. *Science* 268, 681–686.
- Townsend, B.J., Poole, A., Blake, C.J. and Llewellyn, D.J. (2005) Antisense suppression of a (+)- δ -cadinene synthase gene in cotton prevents the induction of this defense response gene during bacterial blight infection but not its constitutive expression. *Plant Physiol.* 138, 516–528.
- Traber, M.G. and Sies, H. (1996) Vitamin E in humans: Demand and delivery. *Annu. Rev. Nutr.* 16, 321–347.
- Venables, B.J., Waggoner, C.A. and Chapman, K.D. (2005) *N*-Acylethanolamines in seeds of selected legumes. *Phytochem.* 66, 1913–1918.
- Venkatramesh, M., Karunanandaa, B.S., Gunter, C., Thorne, G. and Crowley, J.J. (2000) Increased sterol content in oilseeds through biotechnology. *Annual meeting of the American Society of Plant Physiology*, San Diego, CA, July 15–19, 2000.
- Voelker, T. and Kinney, A.J. (2001) Variation in the biosynthesis of seed-storage lipids. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 52, 335–361.
- Voelker, T.A., Hayes, T.R., Cranmer, A.M., Turner, J.C. and Davies, H.M. (1996) Genetic engineering of a quantitative trait: metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. *Plant J.* 9, 229–241.
- Voelker, T.A., Jones, A., Cranmer, A.M., Davies, H.M. and Knutson, D.S. (1997) Broad-range and binary-range acyl-acyl-carrier protein thioesterases suggest an alternative mechanism for medium-chain production in seeds. *Plant Physiol.* 114, 669–677.
- Voelker, T.A., Worrell, A.C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D.J., Radke, S.E. and Davies, H.M. (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science* 257, 72–74.
- Waterhouse, P.M., Graham, M.W. and Wang, M.-B. (1998) Virus resistance and gene silencing in plants is induced by double-stranded RNA. *Proc. Nat. Acad. Sci. USA.* 95, 13959–13964.
- Weselake, R., Shah, S., Taylor, D., Harwood, J., Laroche, A., Moloney, M., Rakow, G. and Raney, P. (2006) Increasing seed oil content in canola through over-expression of type-I diacylglycerol acyltransferase. *17th International Symposium on Plant Lipids*. East Lansing, MI, July 16–21, 2006.
- Whillhit, G. (2003) Oil seed production: U.S. cottonseed industry facing hurdles. *Oil Mill Gazetteer* 18, 4–8.
- Wu, Y.R., Machdo, A.C., White, R.G., Llewellyn, D.J. and Dennis, E.S. (2006) Expression profiling identifies genes expressed early during lint fiber initiation in cotton. *Plant Cell Physiol.* 47, 107–127.
- Zhang, D., Pirtle, I.L., Park, S.J., Nampaisansuk, M., Neogi, P., Wanjie, S.W., Pirtle, R.M., and Chapman, K.D. (2009) Identification and expression of a new delta-12 fatty acid

- desaturase (*FAD2-4*) gene in upland cotton and its functional expression in yeast and *Arabidopsis thaliana* plants. *Plant Physiology et Biochemistry*, in press.
- Zhou, X-R., Singh, S., Liu, Q. and Green, A. (2006) Combined transgenic expression of $\Delta 12$ -desaturase and $\Delta 12$ -epoxygenase in high linoleic acid seeds to increased accumulation of vernolic acid. *Func. Plant Biol.* 33, 585–592.
- Zock, P.L. and Katan, M.B. (1992) Hydrogenation alternatives: effects of *trans* fatty acids and stearic acid versus linoleic acid on serum lipids and lipoproteins in humans. *J. Lipid Res.* 33, 399–410.
- Zarins, Z.M. and Cherry, J.P. (1981) Storage proteins of glandless cottonseed flour. *J. Food Sci.* 46, 1855–1859.

Genomics of Cotton Fiber Secondary Wall Deposition and Cellulose Biogenesis

Candace H. Haigler, Bir Singh, Guirong Wang, and Deshui Zhang

Abstract The deposition of > 90% cellulose in the cotton fiber secondary wall makes this unique cell powerful for understanding cellulose biogenesis, a process with great importance in nature and industry. This chapter provides an overview of cellulose biogenesis, summarizes how cotton fiber has previously facilitated unique insights in this field, and explains how cellulose is important in terms of cotton fiber physical properties. The nature of the cotton fiber secondary wall transcriptome is discussed, including comparisons to primary-wall-stage fiber and the Arabidopsis proteome. Microarray data, including validation by quantitative reverse transcription PCR, are described to show that transcriptomes for secondary wall deposition in cotton fiber and xylem are similar. The functional context of selected genes that are up-regulated for secondary wall deposition is discussed.

1 Introduction

Cellulose is an essential renewable material in nature and industry, and we need to know more about how plants carry out a nanoscale manufacturing process to produce it. Cellulose is defined chemically as polymeric β -1,4-glucan, but it occurs naturally as partly crystalline fibrils within plant cell walls. Therefore, the production of cellulose is best referred to as a “biogenetic” rather than simply a “biosynthetic” process. Cellulose fibrils are typically 3 – 6 nm in diameter and several μ m long; for example, a secondary wall cellulose molecule with 10,000 glucose units would be 4.75 μ m long. The term “microfibril” historically was used to denote the primary unit of cellulose biogenesis in higher plants, but given their diameter and the importance of nanoscale phenomena in their synthesis and interactions with other cell wall molecules, “nanofibril” is

C.H. Haigler (✉)

Department of Crop Science and Department of Plant Biology, North Carolina State University

e-mail: candace_haigler@ncsu.edu

more appropriate (Haigler 2006). Due to their high stiffness and strength, cellulose nanofibril orientation is a primary control point for the plant morphogenetic processes that depends on directional constraint of cell expansion. Thin, extensible, primary walls that typically contain 10 – 20% cellulose surround all plant cells. After cell expansion is largely complete, some cells deposit cellulose-rich secondary walls that are at least 4 μm thick. These confer extra strength to cells such as fibers and water-conducting tracheary elements (Haigler 1985; Niklas 1992; Stone 2001).

Cellulose quantity and physical properties are central to virtually every industrial use of fiber and chemical cellulose. Wood is composed mainly of vascular fibers and water conducting cells with secondary walls, and the cellulose component rather than the lignin is the useful component for many products. Cotton fiber, which is essentially a cellulosic secondary cell wall when mature, is a major resource for the textile and chemical industries. Since secondary walls derive most of their extra strength through modified cellulose properties (Hsieh 1999; Benedict et al. 1999), there is a particular need to understand the regulation of secondary wall cellulose biogenesis as a basis for devising strategies to stabilize and/or change the quantity and quality of cellulose. This need has been emphasized recently by the interest in making biofuels from renewable plant biomass, with the cellulose component being the most abundant and useful for saccharification.

For numerous reasons, cotton fiber provides a valuable experimental tool and genomics system for understanding the control of cellulose biogenesis in plants in general (Kim and Triplett 2001). (1) The normal cotton fiber secondary wall has the highest percentage of cellulose known in plants. Residual weight of crystalline cellulose after solubilization of other components equaled 90% of original fiber weight (Haigler et al. 2007). However, since both the primary wall and non-crystalline cellulose would have been removed by the acetic/nitric solvent, >90% cellulose existed in the secondary wall *per se*. In contrast, typical secondary walls in xylem/wood contain 40–50% cellulose, although bending-induced tension wood has special G-fibers with a secondary wall layer that is richer in cellulose (Pilate et al. 2004). (2) Cotton fibers, after the initiation and earliest elongation stages, are easily separated from the seed in large quantity using minimally stressful procedures, so gene expression in the context of a single cell is easily analyzed. (3) The prolonged (45 – 60 day) developmental program of *Gossypium hirsutum* includes distinct phases of primary wall deposition (for fiber elongation) and secondary wall deposition (for fiber thickening), as well as an intervening transitional stage. (4) Experimentation is made more efficient by year-round fiber development in greenhouses or growth chambers or on cultured ovules. (5) The genus *Gossypium* has a well-documented and fascinating evolutionary history that includes recent polyploidization. Both diploid and allotetraploid species were domesticated to create cultivated crops by several different groups of indigenous people beginning ~6000 years ago, and breeding has continued until the present day (Fryxell 1979; Endrizzi et al. 1985; Adams and Wendel 2004). (6) Genomic tools

for cotton are rapidly emerging, as described further elsewhere in this volume and below.

This chapter will emphasize genomic research in cotton and other plants that provides additional evidence and/or extends concepts described previously. From a pre-genomic perspective, the fundamental biology of cotton fiber cellulose was reviewed recently including: (a) the role and changing properties of cellulose during fiber development; (b) control of cellulose biogenesis by genetic, cellular, and biochemical processes; and (c) limitations on cellulose accumulation imposed by abiotic stresses (Haigler 2007). In the brief summary below, primary citations are often not repeated if they can be found there or in other reviews on cellulose biogenesis (Delmer 1999; Haigler et al. 2001; Scheible and Pauly 2004; Saxena and Brown 2005; Somerville 2006). In addition, a recent multi-author book provides a good resource for understanding diverse aspects of cellulose biogenesis (Brown and Saxena 2007), and other reviews give more comprehensive treatment of several aspects of cell wall biosynthesis discussed here (e.g. Farrokhi et al. 2006).

2 Overview of Cellulose Biogenesis

Cellulose biogenesis is under complex, multi-level, control including regulation through genetics, cell biology, biophysics, biochemistry, hormones, and environmental interaction. This speaks to both the nanoscale cellular manufacturing of crystalline fibrils and the advantage to the plant of regulating carefully the strong, irreversible, carbon sink of secondary wall cellulose deposition. Cellulose biogenesis bridges both sides of the plasma membrane, from cortical cytoskeleton to cell wall space. Plant cellulose fibrils are synthesized by a large organized complex of proteins, which has been detected as a “rosette” of 6, ~8 nm diameter, particles in the inner face of the fractured plasma membrane. The rosette marks the existence of protein transmembrane domains within an orderly aggregation of ~24–36 cellulose synthase (CESA) enzymes, and probably other proteins that participate in cellulose synthesis. Microfibril-terminal-complex (TC) or “rosette” are traditional names for this complex, but “cellulose synthase complex (CSC)” has been suggested as an alternative (Diotallevi and Mulder 2007). However, CSC to denote “cellulose *synthesis* complex” may be preferred to allow the possibility that other proteins besides CESAs are part of the complex; this terminology will be used further in this article.

The close proximity of synthetic enzymes facilitates self-assembly and co-crystallization of newly polymerized β -1,4-glucan chains. At least the core part of the CSC is transported to the plasma membrane by fusion of Golgi vesicles, and CSC integrity in the plasma membrane may be under oxidative and proteolytic regulation. Supporting a long-standing hypothesis, there are now both *in vivo* and biophysical modeling evidence that CSCs move at about 6×10^{-9} m/s through the fluid plasma membrane as cellulose nanofibrils

crystallize behind them (Paradez et al. 2006; Diotallevi and Mulder, 2007). Microtubules and the cellulose nanofibrils themselves are seemingly interdependent in determining the localization and orientation of cellulose fibril biogenesis, at least in some cell types. A plasma membrane- and/or cytoskeleton-associated form of sucrose synthase participates in secondary wall cellulose biogenesis by cleaving sucrose to release UDP-glucose (UDP-glc) to the polymerizing enzymes. Metabolic control, interfacing with hormonal and environmental regulation, is required to regulate the extent of cellulose deposition in different cell types and in response to different growing conditions. In reference to cotton genes that are up-regulated for secondary wall deposition, we will discuss some proteins that participate in this process in more detail below. However, it is important to note that there is currently a poor understanding of exact mechanisms of protein action in most cases or how diverse participants carry out the integrated process of cellulose biogenesis.

3 Cotton Fiber Illuminates the Control of Cellulose Biogenesis

Although extensive cell wall thickening is characteristic of the seed hairs of all *Gossypium* species (Fryxell 1979), unless specifically noted otherwise, research data summarized here are for *Gossypium hirsutum*. Allotetraploid *G. hirsutum* (AADD genome) arose 1–2 mya by hybridization of two diploid genomes, AA (typified by *G. arboreum*) and DD (typified by *G. raimondii*) (Wendel and Cronn 2002). Domesticated *G. hirsutum* is now grown most commonly for fiber crop production worldwide, and *G. arboreum* is sometimes grown on a much smaller scale in Asia. However, *G. raimondii* produces short fiber that is not economically useful, although a similar DD genome contributed useful quantitative trait loci (QTL) for fiber traits to the allotetraploid (Jiang et al. 1998). Timing of developmental events in cotton fiber is referenced as days post anthesis (DPA) because fiber initiation occurs at or near anthesis (flower opening). Typical primary walls, ~ 0.5 μm thick and containing 20–25% cellulose along with pectin, xyloglucan, and protein, are synthesized during fiber elongation. In *G. hirsutum*, primary wall deposition proceeds alone until 15–19 DPA. Next, a transition phase with concurrent primary and secondary wall deposition begins within the range of 16–20 DPA. At this stage the wall begins to thicken rapidly, including synthesis of cellulose fibrils that are oriented at 45–55° relative to the long fiber axis within the transiently-synthesized “winding” cell wall layer (Ryser 1985). Beginning within the range of 19–24 DPA, continued fiber wall thickening occurs via $> 90\%$ cellulose deposition until 45–50 DPA when the boll opens and the fiber dries (Meinert and Delmer 1977). The ranges of DPA for developmental stages shown above reflect variability within cultivars and across different environments.

Because of the distinct phase of secondary wall deposition within its prolonged development program, cotton fiber has a long history of revealing

new information about the control of cellulose biogenesis and secondary wall deposition. For example, research on cotton fiber: (1) allowed the first plant *CESA* genes to be identified (Pear et al. 1996); (2) illuminated effects of microtubules and actin on the orientation of cellulose fibrils (Seagull 1993); (3) showed for the first time that multiple tubulin isotypes existed in a single cell, with some isotypes under developmental regulation (Whittaker and Triplett 1999); (4) showed changing rates of cellulose biogenesis and changing length of cellulose chains for primary and secondary wall deposition (Meinert and Delmer 1977; Timpa and Triplett 1993); (5) implicated UDP-glc as the substrate for cellulose biogenesis (Franz 1969; Carpita and Delmer 1981); (6) demonstrated a role for membrane-associated sucrose synthase acting degradatively in supplying UDP-glc for secondary wall cellulose biogenesis (Amor et al. 1995; Salnikov et al. 2003); (7) illuminated the cool temperature sensitivity of cellulose biogenesis, the associated block in intra-fiber sucrose synthesis, and the ability of additional sucrose supply to boost fiber filling under stressful conditions (Gipson 1986; Roberts et al. 1992; Martin and Haigler 2004; Haigler et al. 2007); (8) revealed a role for an oxidative burst, probably mediated by a small GTPase Rac, in initiating secondary wall deposition (Potikha et al. 1999); (9) showed that oxidation can promote *CESA* interaction via their N-terminal zinc-binding domain, and thereby probably affect CSC stability (Kurek et al. 2002); and (10) showed that, unusually for a membrane protein, *CESA* proteins have a short half-life in the plasma membrane (≤ 30 min), and that their turnover and degradation are likely mediated through activity of a metallothionein protein and proteolysis (Jacob-Wilk et al. 2006). Other tantalizing clues about cellulose biogenesis obtained from cotton fiber are in need of further exploration, e.g. lipid intermediate that may exist (Peng et al. 2002, although for critique see Schrick et al. 2004). These past results lead to the expectation that further research on cotton fiber will be useful not only for cotton improvement, but also for revealing novel mechanisms controlling secondary wall deposition.

4 The Relationship of Cellulose to Fiber Physical Properties

A survey of the impact of cellulose on cotton fiber utility makes it clear why genomic analysis to identify underlying control mechanisms for cellulose properties is so important for our future potential to improve the usefulness of cotton fiber through breeding complemented with genetic engineering or marker assisted selection. In contrast to being a bulk commodity, future cotton crops may have special fiber and cellulose properties that are directed toward particular high value products. However, achieving this goal requires complete understanding of underlying mechanisms governing all aspects of fiber differentiation.

A key character of technologically useful cotton fiber is an intermediate thickness (maturity) of the secondary wall relative to the fiber diameter. Fiber

must be neither immature (see below) nor overly mature because the lumen (former vacuolar space) must remain large enough to allow collapse into an ellipsoid/ribbon-like shape and development of spiral twists (convolutions) that aid spinning into yarn (Hutchinson et al. 1945). A fiber that is over filled with secondary wall remains cylindrical and unconvoluted upon drying (see diagram in Haigler et al. 2005). Modern cotton typically has 3.5 – 6.9 convolutions/mm fiber, with more twists increasing single fiber strength (Hsieh 1999). Convoluted fiber is characteristic of and considered diagnostic of human domestication of archaeological cotton in the New World, first to make twined textiles and nets then woven cloth (Stephens 1970). The desirable degree of secondary wall thickening increases fiber mass (yield) along with single fiber strength, dyeing intensity, and water absorption. The effects of increased secondary wall thickness on single fiber strength, Micronaire, and Maturity Ratio are illustrated by a family of transgenic cotton plants over-expressing sucrose phosphate synthase (Haigler et al. 2007).

In addition, molecular properties of cellulose, including degree of polymerization (DP), fibril orientation, percent crystallinity, and crystallite size affect fiber quality characteristics of modern cultivars as summarized briefly here and reviewed thoroughly elsewhere (Lewin and Pierce 1998; Benedict et al. 1999; Hsieh 1999; Bradow and Davidonis 2000). Percent crystallinity, meaning the percentage of cellulose in crystalline vs. disordered state (e.g. on fibril surfaces), has been positively correlated with fiber bundle strength and single fiber strength in some studies (Hindeleh et al. 1980; Hsieh et al. 1997), but not others (Timpa and Ramey 1994). These relationships may be variety dependent, and comparing different studies is especially problematic (Hsieh et al. 1999). Percent crystallinity is also maximized in correlation with more cellulose fibril bundling in the fiber secondary wall (Willison and Brown 1977) where cellulose can self-assemble without substantial interference from other molecules (Haigler 1991).

Cellulose crystallite size, meaning the average width of single crystals, varies with developmental stage. Cotton primary wall cellulose exists as cellulose IV due to poor order in the lateral dimension, indicating small fibril diameter and crystallite size (Chanzy et al. 1978). Cotton secondary walls contain more orderly cellulose I fibrils with higher crystallite size (3.5 – 5.5 nm) (Ryser 1999; Hu and Hsieh 1996). Crystallite size is mediated by the number of β -1,4 glucan chains synthesized in close proximity and by matrix molecules that can interfere with cellulose self-assembly (Haigler 1991). Cellulose crystallite size is positively correlated with single fiber breaking force (Hsieh et al. 1997), but negatively correlated with water and dye absorption. The latter negative effect arises because the crystalline core of the cellulose nanofibril cannot be penetrated by water or dyes (Rowland and Bertoniere 1985).

The cellulose degree of polymerization (DP_w, weight averaged) in cotton fiber primary walls was lower (~4000) than in secondary walls (~11,000), with distribution around the mean occurring in each case (Timpa 1991; Timpa and Triplett 1993). The average cellulose DP in solubilized whole fiber and in

acetic/nitric acid-resistant crystalline regions were positively correlated with bundle strength (Timpa and Ramey 1989, 1994; Benedict et al. 1994). The DP distribution has its own importance for fiber quality, greatly impacting flexibility and processing potential (Schneider et al. 1996), and it also affects manufacturing of cellulose derivatives (Heinze 1998).

More highly oriented cellulose fibrils increase fiber strength (Yatsu 1983), and a small fibril angle relative to the longitudinal fiber axis correlates with higher elastic modulus (implying less extensibility or extension to break) and strength. Fibril angle is also correlated with resilience, work, recovery, and permanent set (Rebenfield 1990). There is a developmental shift in the orientation of cellulose fibrils relative to the longitudinal fiber axis, with angles of $\sim 70^\circ$ - 90° , 45° - 55° , and 20° during the primary wall, transition, and secondary wall stages, respectively (Arthur 1990; Rebenfield 1990). The mechanisms mediating these changes and interrelating cellulose fibrils and microtubules that shift in parallel are subjects of ongoing research (Baskin 2001).

5 Analysis of the Cotton Fiber Secondary Wall Transcriptome

5.1 Specialized cDNA Library Production and EST Sequencing

In order to complement prior sequencing of genes expressed at the primary wall stage of fiber development (Arpat et al. 2004; <http://demeter.bio.bnl.gov/>), a set of *G. hirsutum* fiber secondary wall (G.h.fbr-sw ESTs) was generated at Michigan State University from a suppression subtraction hybridization (SSH) library (Diatchenko et al. 1996). By use of 20 DPA fiber of *G. hirsutum* cv. Deltapine90 as tester and 6 DPA fiber as driver, clones that were up-regulated at 20 DPA compared to 6 DPA were expected to be enriched, and 9,240 high quality sequences were deposited into GenBank with accession numbers C0490611-C0499850 (Haigler et al. 2005). Both analysis of wall thickness in the fiber used to make the library and subsequent analysis of gene expression patterns (see below) indicated that 20 DPA was early in the primary- to secondary- wall transition period for fiber developing under relatively cool temperature in a Michigan greenhouse (J. Klug, personal communication) or subsequently in a tightly controlled 26/22°C greenhouse at North Carolina State University (U. Avci and C. Haigler, unpublished). During the transition, gene expression required for primary and secondary wall deposition is expected to be decreasing and increasing, respectively.

By reference to a comprehensive assembly of 150,436 cotton ESTs (Udall et al. 2006; http://agcol.arizona.edu/cgi-bin/pave/adv_search.cgi?AssemblyID=Cotton12), we analyzed the new gene discovery rate arising from three strategies for making cotton fiber cDNA libraries (Table 1). (Note that the G.h.fbr-sw ESTs are called GH_SCW in the cotton comprehensive gene assembly.)

Table 1 Analysis of individual cotton fiber cDNA libraries in relationship to the cotton comprehensive assembly (Udall et al. 2006)

Fiber EST data set	Genotype and fiber stages	Library ESTs in the assembly	Unigenes with library ESTs	Average ESTs/unigene ^a	Rate of gene discovery (%) ^b
GA_Ea	<i>G. arboreum</i> cv AKA8401; 7–10 DPA fiber	28,080 (of 46,603 in GenBank)	11,816	2.38	42.08
GH_BNL	<i>G. hirsutum</i> cv Acala Maxxa; 5 DPA fiber	7,590 (of 8,910 in GenBank)	4,798	1.58	63.21
G.h.fbr-sw a.k.a. GH_SCW	<i>G. hirsutum</i> cv DP 90; 20 DPA fiber	7,372 (of 9,240 in GenBank)	3,253	2.27	44.13

^a Calculated as (library ESTs included in the assembly/unigenes containing library ESTs).

^b Calculated inversely as (unigenes containing library ESTs/library ESTs included in the assembly) \times 100.

The first generation cotton assembly contained 17,355 tentative unigenes that included ESTs from at least one of the three fiber-specific libraries. The *G. arboreum* GA_Ea ESTs were generated from a standard cDNA library representing 7 – 10 DPA fiber, but during sequencing two successive rounds of filter colony hybridization were used to remove the most abundant clones already sequenced (Arpat et al. 2004). The *G. hirsutum* GH_BNL ESTs were generated from a cDNA library normalized by use of reassociation kinetics hybridization (Bonaldo et al. 1996) using RNA isolated from 5 DPA fiber (B. Burr, personal communication and <http://demeter.bio.bnl.gov/>). Normalization is essentially a self-subtraction process using the same cDNA population as both tester and driver in order to create similar representation of all transcripts regardless of their original abundance in cells.

As shown in Table 1, the GH_BNL sequence set was the least redundant, although redundancy would have been inevitably increased through the much deeper GA_Ea sequencing. In another analysis in which assembly unigenes with 1, 2, or 3 ESTs from any one library were considered together as low redundancy sequences, 75% of the unigenes including GH_BNL ESTs had low redundancy compared to 52% and 50% of unigenes including GH_SCW or GA_Ea ESTs, respectively (data not shown). Therefore, in terms of new gene discovery, sequencing of the GH_BNL normalized fiber library was most cost effective, as has also been observed in other species (Zhang et al. 2004a). The SSH cDNA library construction method that combined subtraction and normalization did not aim to enhance the rate of gene discovery *per se*; rather it helped to enrich differentially expressed genes within the G.h.fbr-sw/GH_SCW ESTs. These outcomes were confirmed by comparison to the Arabidopsis proteome (see below).

5.2 Cotton Fiber Gene Expression Compared to *Arabidopsis*

The proteome of the model plant *Arabidopsis thaliana* was used as a reference point for analyzing the changing complexity of gene expression during fiber development. Unigenes from the GA_Ea, GH_BNL, and GH_SCW libraries, which had been subjected to uniform length and quality editing, were downloaded from the cotton comprehensive gene assembly (numbers in Table 1). Then each unigene set was compared by blastx to the *Arabidopsis* proteome with an E-10 cut-off value as used previously to indicate significantly similar plant sequences (Gutierrez et al. 2004). The number of unique *Arabidopsis thaliana* gene identifiers (At Gene IDs, found at TAIR, <http://www.arabidopsis.org>) matched by each unigene set were: (a) 5,682 (48.09% of unigenes) for the GA_Ea library; (b) 3,149 (65.63% of unigenes) for the GH_BNL library; and (c) 1,770 (54.44% of unigenes) for the GH_SCW library. The inclusion of both A genome *G. arboreum* and AD genome *G. hirsutum* in such a conceptual analysis is justified by the relatively recent polyploidization event and by the fact that both species produce commercially useful fiber that develops similarly. Note that a large set of *G. hirsutum* sequences (11,692 unigenes) recently deposited into GenBank derived from ovules with attached fiber (see p. 654 in Shi et al. 2006), so these sequences were not included in the results shown here. However, when this sequence set was included in the same type of analysis of *G. hirsutum* sequences only, it does not change the conclusion regarding unique gene expression at the secondary wall stage of fiber development (see below).

Comparison of cotton fibers ESTs to the *Arabidopsis* proteome showed that a single cell is not a simple cell in terms of its set of expressed genes. In the cotton comprehensive gene assembly, unigenes with contributions from GA_Ea, GH_BNL, and G.h.fbr-sw/GH_SCW fiber libraries were classified as including ESTs from the: (1) primary wall stage; (2) secondary wall stage; or (3) both stages of fiber development. Each group of translated fiber-associated unigenes was then compared by blastx to the *Arabidopsis* proteome (E-10 cut-off value), and each list was made non-redundant in terms of At Gene IDs recognized. The outcomes were used to generate a Venn diagram illustrating current knowledge about fiber-specific gene expression complexity (Fig. 1). The 7,229 total non-redundant At Gene IDs represent 36.4% of 19,867 cotton fiber-associated unigenes analyzed. Interestingly, in Group 1 of primary wall genes, 1,016 of 6,698 At Gene IDs arose from the GH_BNL ESTs, which shows that normalization-based sequencing from *G. hirsutum* was useful in complementing the large scale sequencing effort from *G. arboreum*.

Of 6,698 At Gene IDs matched by primary wall sequences and 1,770 matched by secondary wall sequences, 1,239 At Gene IDs were shared between the stages of development. At the onset of secondary wall deposition, cotton genes were expressed that were homologous to 531 unique At Gene IDs, representing 7.3% of the total. This reflects a minimum estimate of gene expression for secondary wall deposition because only 19.6% of the unigenes

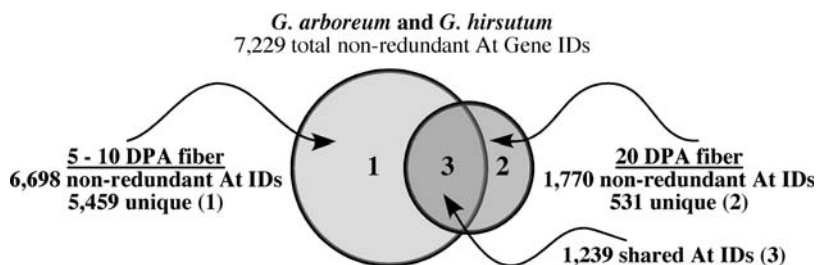


Fig. 1 Venn Diagram illustrating genetic complexity during primary and secondary wall deposition in cotton fiber development in reference to homology ($\leq E-10$) of translated cotton unigenes to the Arabidopsis proteome. The 5 – 10 DPA group was based on cotton unigenes with contributions from GA_Ea and/or GH_BNL sequences, whereas the 20 DPA group derived from only GH_SCW unigenes

in the three fiber libraries arose from sequencing at the secondary wall stage of fiber development. Remembering that cotton genes without close Arabidopsis homologues are not represented in this analysis, these results demonstrate that genes equivalent to at least 25 – 30% of the proteome of the whole Arabidopsis plant are expressed during cotton fiber development.

Previously, G.h.fbr-sw sequences were assigned to MIPS major categories of protein function by analogy to homologous Arabidopsis proteins. Highly represented categories were consistent with control of cellular organization and transport processes, as well as metabolic processes, in transition stage cotton fiber. Cotton sequences related to these functions were also over-represented compared to the Arabidopsis proteome, whereas transcriptional functions were under-represented (Haigler et al. 2005). A similar analysis was done more recently for GO functional categories, along with comparison of categories for cotton genes up-regulated on the microarray to be discussed later (Fig. 2). In the Molecular Function category, high enzymatic activity as well as transport activity was indicated for transition-stage cotton fiber (hatched bars), whereas transcriptional activity was implicated to a lesser extent than for the entire Arabidopsis proteome (gray bars). The latter difference is consistent with the more restricted set of developmental changes occurring at one stage in a single cell compared to a whole plant. However, regulatory activity via kinases and, potentially, protein binding is well represented within the G.h.fbr-sw sequences. Other analyses show that substantial numbers of the hydrolases and transferase are related to sugar- and/or cell wall metabolism (data not shown), which is consistent with dynamic changes in the fiber cell wall occurring at the transition stage. In the Cellular Components category, most G.h.fbr-sw sequences were in the “other” categories just as in Arabidopsis. However, compared to Arabidopsis, there was an enhancement of the percentage of G.h.fbr-sw sequences in the categories of ER, Golgi, cell wall and plasma membrane, which makes sense given the main activities of the secondary-wall-stage cotton fiber. In the Biological Processes category, G.h.fbr-sw sequences

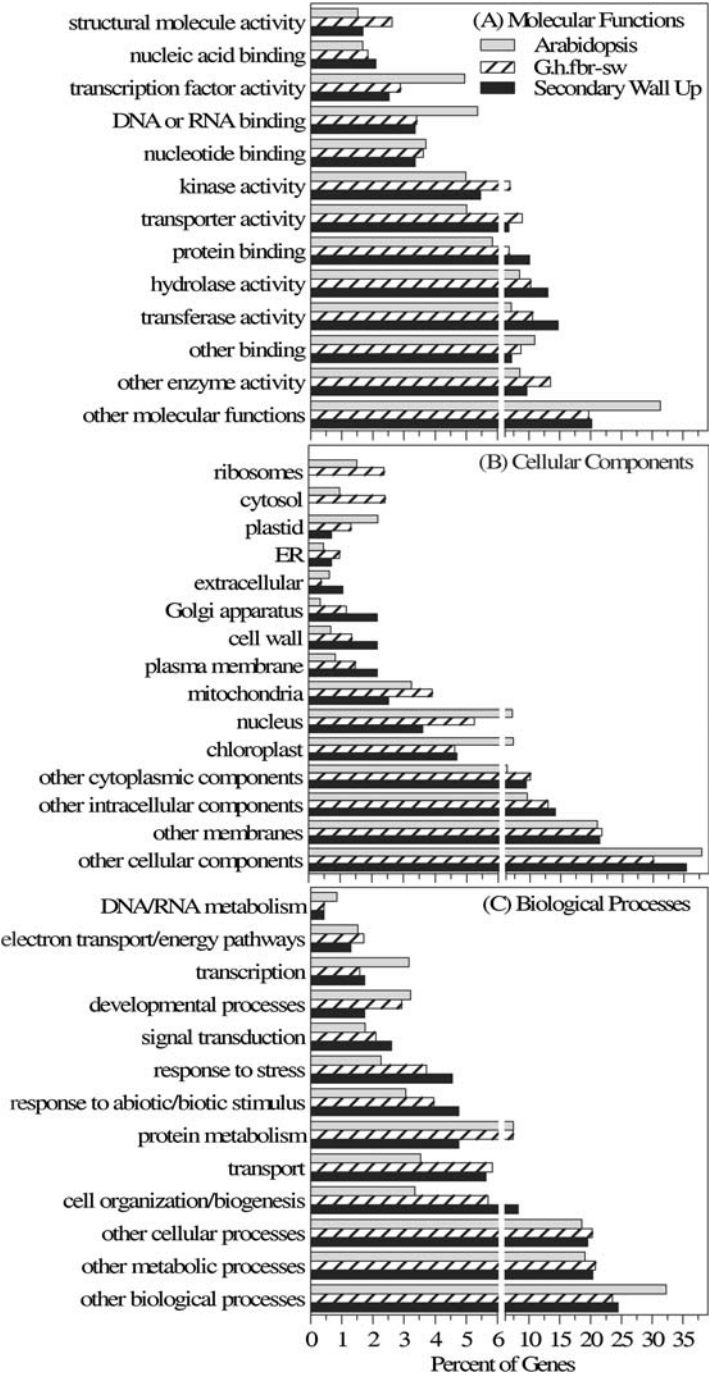


Fig. 2 (continued)

were again highly represented in the “other” categories, but there was over-representation compared to Arabidopsis in the categories of transport and cell organization and biogenesis.

5.3 A Discrete Transcriptome for Secondary Wall Deposition

Microarray analysis was done by Curtis Wilkerson and Jeff Landgraf at the Research Technology Support Facility, Michigan State University. Spotted onto slides in duplicate were 3,185 cDNA clones after PCR amplification; these represented 93.1% of 3,420 internally-assembled unigenes. (Note that 167 fewer GH_SCW unigenes were included in the cotton comprehensive gene assembly because of preliminary uniform editing of sequence length and quality between all sequence sets included.) RNA used to make probes was isolated from fiber at 6 and 10 DPA (elongation via primary wall deposition), 20 DPA (transition to secondary wall deposition under relatively cool conditions), and 24 DPA (onset of high rate secondary wall cellulose biogenesis). The experiment was designed to reveal genes that were up-regulated for secondary wall deposition. To exemplify the sensitivity of the array, for clones up-regulated at 20 vs. 10 DPA, the minimum significant change ($p \leq 0.050$ on combined duplicates of each cDNA spotted onto each slide) was $\log_2 = 0.5$, or 1.4-fold, but 84.8% of significant changes had magnitudes $\geq \log_2 = 1$, or 2-fold.

General trends in the microarray data are further described here in terms of clones that were significantly up-regulated at 20 DPA vs. 6 DPA and 10 DPA ($p \leq 0.050$ on combined duplicates). This is a stringent requirement to enhance the chances that real changes relative to onset of secondary wall deposition are discussed. There were 276 clones that were up-regulated at the transition to secondary wall synthesis vs. 96 with enhanced expression during primary wall synthesis (Fig. 3). The existence of 2.9 times more clones in the up-regulated category reflects the initial bias of the spotted cDNAs toward the secondary wall stage of fiber development. Graphing the expression patterns of genes with changes at 20 DPA shows, as expected, that the discrete developmental stages of primary and secondary wall deposition in cotton fiber are underpinned by substantially different programs of gene transcription. The 276 up-regulated clones, when translated *in silico* and compared to the Arabidopsis



Fig. 2 (continued) GO Annotation Groups (A, B, C) for the G.h.fbr-sw sequence set (striped), for cotton genes up-regulated at 20 DPA on the microarray (black), and for the Arabidopsis proteome (grey). The scale is expanded for 0–6%. Cotton genes were grouped in analogy with their homologous Arabidopsis gene ($\leq E-10$; <http://www.arabidopsis.org/tools/bulk/go/index.jsp>). Percentages are based on the entire Arabidopsis proteome or 1770 or 181 unique At Gene IDs corresponding to G.h.fbr-sw sequences or its subset that was up-regulated at 20 DPA, respectively

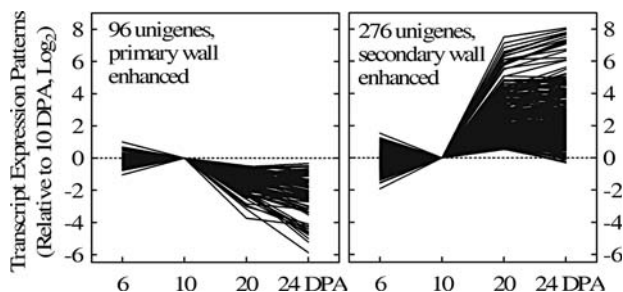


Fig. 3 Graphical summary of microarray results for genes with significantly different expression at 20 DPA compared to 6 and 10 DPA ($p \leq 0.05$). The expression pattern of each gene is shown relative to its own level at 10 DPA. The graphs show that developmental shift between primary and secondary wall deposition (around 20 DPA in this experiment) has a strong transcriptional basis. Rigorous methods were used to generate these results, including two copies of each cDNA spotted on each slide and 3 biological replications and two technical replications (dye swaps) for each data point, followed by quality control then statistical analysis of the results (see also R Development Core Team, 2005; Smyth 2004, 2005; Smyth and Speed 2003; Smyth et al. 2005). Briefly, RNA was labeled with the amino-allyl procedure (<http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml>). Hybridization was performed using SlideHyb #1 buffer (Ambion). Slides were scanned using an Affymetrix 428 ArrayScanner and analyzed using the GenePix Pro 3.0 software (Axon). Array normalization and statistical analysis were performed using the “limma: Linear Models for Microarray Data” library module (version 2.2.0) of the R statistical package (version 2.2.0). Slide intensity data were normalized using the global loess method. The least squares method was used for the linear model fit utilizing the Benjamini and Hochberg method to control the false discovery rate

proteome by BLAST, were homologous ($E \leq -10$) to 181 unique Arabidopsis proteins. (Accounting for the total of 276 up-regulated clones, 47 did not match Arabidopsis sequences at this cut-off level and 48 matched 23 At Gene IDs redundantly). Of the 181 unique Arabidopsis homologs of the up-regulated genes, only 23 were held in common with genes up-regulated at 24 DPA in *G. arboreum* within the GA_Ea sequences derived from the primary wall stage of fiber development (Arpat et al. 2004). The GO categories of the up-regulated G.h.fbr-sw genes were analyzed according to GO annotations of the Arabidopsis homologs (Fig. 2, black bars). In the Molecular Functions group, categories that were over-represented compared to the entire G.h.fbr-sw sequence set were protein-binding, hydrolase, and transferase. In the Cellular Components group, over-represented categories included Golgi apparatus, cell wall, and plasma membrane. In the Biological Processes group, genes up-regulated at the onset of secondary wall deposition were mainly proportional to their overall representation in the G.h.fbr-sw sequences. These trends are consistent with emphasis on cell wall synthesis and remodeling during the primary-to-secondary wall transition in cotton fiber development.

It was important to validate microarray data by real-time reverse-transcription PCR (qPCR). To establish a basis for this between 6 – 30 DPA in cotton fiber, three possible endogenous controls, *EIF-5*, *Tua-4*, and *UBQ*, were tested

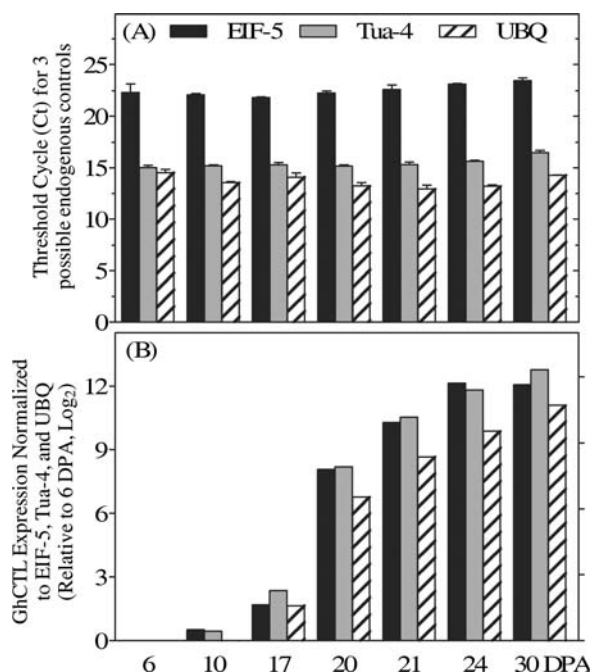


Fig. 4 Comparison of endogenous controls for qPCR analysis of cotton fiber gene expression between 6 – 30 DPA. (A) Expression pattern of three potential endogenous controls, EIF-5 (eukaryotic translation initiation factor-5, Accession: CO492947), GhTua-4 (alpha-tubulin-4, Accession: AY345605), and GhUBQ (ubiquitin, Accession: AY189972). (B) GhCTL expression normalized to each control. Forward (F) and Reverse (R) primer pairs (5' to 3') were: (a) for EIF-5, (F) GGTGTGCCATTGTGCAAGGA, (R) CCGTAGGTGAGCGTTAATCAGA; (b) for GhTua-4, (F) GCGCTGAGTCTGGTGAAG (R) CAGTAGCAAACGGATAACGGTACA; and (c) for GhUBQ, (F) AAGCTCGCAGTGCTCCAGTT (R) GACACTCTTTCCTCAGCCTCTGA. Total RNA was isolated from ≤ 100 mg cotton fiber ground under liquid nitrogen (Spectrum Plant Total RNA Kit, Sigma), purified (On Column DNase Digestion Kit, Sigma), and reverse-transcribed (~ 800 ng/40 μ l TaqMan Reagents, Applied Biosystems). The 25 μ l PCR reactions (with 12.5 μ l SYBR green, 150 nM (F) and (R) primers, 10 ng cDNA) were run [50°C, 2 min; 95°C, 10 min; followed by 40 cycles of (95°C, 20 sec and 60°C, 1 min)] on a Real Time PCR Model 7300 (Applied Biosystems). Dissociation curves were analyzed to verify the absence of non-specific amplification and primer-dimer formation. SDS software was used for automatic analysis of threshold cycle (Ct) values. Each data point derived from three biological samples, each containing fiber from 2 – 10 bolls. For EIF5 and Tua4, the entire assay was repeated twice starting from RNA isolation, and similar results were observed in other experiments for UBQ. Bars = SE values

(Fig. 4). The most variability of expression was observed for *UBQ*, and its ΔR_n vs. Cycle (log graph, where ΔR_n is the intensity of dye fluorescence as a function of cycle number) showed amplification before 10 cycles (data not shown). The difference between the controls is clarified by examining the expression of a cotton chitinase-like gene, *GhCTL1*, normalized to each one.

GhCTL1 was used for this purpose because of background data showing that it was secondary-wall-specific in several types of cotton cells, including fiber (Zhang et al. 2004b). Normalization to *UBQ*, with its early cycle amplification, often resulted in lower values for *GhCTL1* expression compared to the other two controls. The expression of *Tua-4* dropped at 30 DPA, so, when it was used as a normalizer, the apparent expression of *GhCTL1* increased compared to 24 DPA. On a linear scale, the enhancement in *GhCTL1* expression relative to 6 DPA was 4,493-fold at 24 DPA and 4,299-fold at 30 DPA when normalized to *EIF-5*. The corresponding values after normalization to *Tua-4* were 3,584-fold at 24 DPA and 6,916-fold at 30 DPA. Overall, normalization with *EIF-5* resulted in a *GhCTL1* expression pattern that was more consistent with results of promoter-GUS assay and Northern blotting; the latter showed approximately equal expression for *GhCTL1* during the later stages of secondary wall deposition (Zhang et al. 2004b).

To verify the expression patterns for selected cotton sequences showing changes on the array, qPCR was performed (Table 2 and Fig. 5). Information from Arabidopsis was used to aid primer design since the cotton genome is

Table 2 Cotton cDNAs for which microarray gene expression patterns were checked by qPCR. E-values are for the cDNA spotted on the array, and longer assembled cotton unigenes with better homology to the Arabidopsis protein may exist

GenBank Accession Number	Arabidopsis homolog	E-value	Abbreviated Name/ Annotation
Up-regulated on the microarray			
CO493183	At3g16920	1E-39	GhCTL1/2; chitinase-like,
CO498265	At4g18780 CesA8	6E-94	GhCesA1; cellulose synthase
CO497506	At3g43190 SUS4	1E-111	GhSUS; sucrose synthase
CO491535	At1g71880 SUC1	1E-10	GhSUC; sucrose transporter
Down-regulated on the microarray			
CO495014	At5g05270	1E-60	GhCHI; chalcone-flavanone-isomerase-like
CO496859	At2g39700 EXPA4	3E-57	GhEXP; expansin
CO492432	At1g19840	2E-12	GhARP; auxin-responsive-SAUR-protein-like
CO499674	At1g26690	1E-34	GhTTP; transmembrane trafficking protein-like
Unchanged on the microarray			
CO492429	At5g20280 SPS1F	3E-54	GhSPS; sucrose-phosphate synthase
CO498005	At5g17310	2E-49	GhUDPG _P ; UDP-glucose pyrophosphorylase

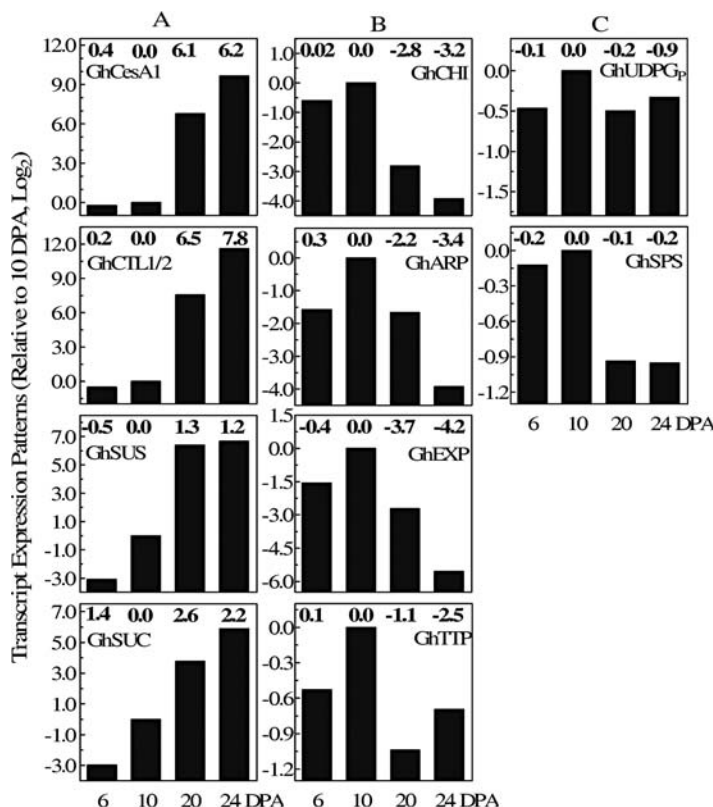


Fig. 5 Comparison between gene expression patterns relative to 10 DPA derived from microarray and qPCR analyses. Columns A, B, and C show genes that were up- or down-regulated or unchanged on the microarray, respectively, as shown by log₂ values in bold above each corresponding ratio in the qPCR histogram. qPCR data (from three biological replicates) were normalized to EIF-5. See Table 2 for accession numbers and Fig. 4 for basic methods. Forward (F) and Reverse (R) primer pairs (5' – 3') were: (a) for *GhCTLI/2*, (F) CCGACCAAGAACGACACGTT, (R) ACCTCGCCACAACTTGAT; (b) for *GhCesA1*, (F) GCAGCAGACGATACAG AATTTCG, (R) CGTTGTTGATTGCGTCT-GAAAC; (c) for *GhSUS*, (F) ACATGG CAGATTACTCCGAGAA, (R) TCCATGCATGCAACATTTC; (d) for *GhSUC*, (F) AGCCATTCCGGTTGAGTTGA, (R) CCAGCCTTTGTATTGGGTCTA; (e) for *GhCHI*, (F) AGCCTGGGAACGTGC-TAGCT, (R) TTGGGAACCTTTGATCT CTTACT; (f) for *GhEXP*, (F) TTTGAC-CAAACCTGCCAATGG, (R) CAGAACT GGCAGTCAAATGCA; (g) for *GhARP*, (F) GCCAGCAACAACACCATCAA, (R) TGAGACTTGTGCCACACAGA; (h) for *GhTTP*, (F) CCCAAAAGCAAGCCA TGTA, (R) TTCATATGGGAATAATTT-CACTCG; (i) for *GhSPS*, (F) GTCTCATGTTCTTGGTGAGCAAT, (R) CACCAGATAGAAGAGCA GCAGAGT; (j) for *GhUDPG_p*, (F) CAACTTTTCTCTGGTTTTGCA, (R) TCAAGACAA-TCCCAAGCATCAT

not completely sequenced. This strategy was justified in light of the close relationship between the two species: *Arabidopsis* (Brassicales) and *Gossypium* (Malvales) are constituents of Rosids II, a large subclade within Rosids in the plant phylogenetic tree (Soltis et al. 2005). Cotton contigs/ESTs were analyzed via blastx to identify the set of closest *Arabidopsis* protein homologs. In order to minimize the chances of amplification of multiple related cotton genes by avoiding highly conserved domains, primers were designed within regions of the cotton sequence that showed divergence from the most closely related *Arabidopsis* sequences. Primer Express 3.0 software (Applied Biosystems) was used to design forward and reverse primers for an amplicon size between 50–150 base pairs. Primers were analyzed by BLAST to show that they only matched one known cotton unigene (or sometimes two closely related sequences).

Figure 5 shows a comparison of gene expression profiles obtained by qPCR (histograms) and microarray analysis (in bold type over each histogram bar for the same DPA). Importantly, for several up- and down-regulated genes, the qPCR data demonstrate that 20 DPA was transitional between the gene expression levels at 10 DPA (high-rate elongation) vs. 24 DPA (high-rate secondary wall deposition). On 20 DPA, expression of genes putatively related to auxin response, which is most important during fiber elongation, and primary cell wall flexibility via expansin action were declining, whereas expression of a secondary-wall-specific isoform of CesA was increasing (see also Guo et al. 2007). Corresponding to its higher dynamic range (Busch and Lohmann 2007), qPCR often revealed greater magnitudes of up/down-regulation than microarray analysis. In a large percentage of cases, the same pattern of gene expression was revealed by the two techniques. However, GhSPS showed apparently constant expression on the microarray, whereas the isoform assayed by qPCR was down-regulated at the onset of secondary wall deposition. Spotted cDNAs may hybridize with multiple closely related genes (Busch and Lohmann 2007), whereas, in general, primers for qPCR are more discriminatory. However, this cannot be fully evaluated for any organism with an unsequenced genome.

6 Cotton Fiber Secondary Wall Transcriptome vs. Xylem

We aimed to allow functional information derived in *Arabidopsis* to illuminate the possible significance of homologous gene expression in cotton fiber. The most abundant secondary walls in plant vegetative tissue occur in xylem, and xylem secondary walls in dicotyledonous plants contain 40 – 50% cellulose as well as substantial amounts of xylan and lignin. As supported by the gene expression data described below, secondary wall deposition in cotton fiber and in water-conducting xylem cells shares common elements. This similarity is reasonable since formation of xylem secondary walls in vascular tissue was a

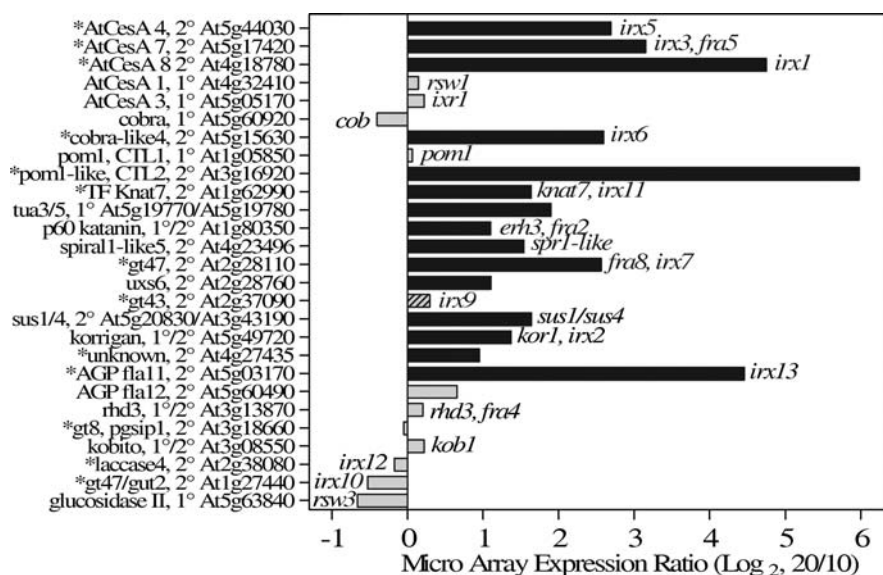


Fig. 6 Microarray results (20/10 DPA expression ratio) for cotton homologs of Arabidopsis cell-wall-associated genes. Black bars indicate significant changes ($p \leq 0.05$), and the striped bar indicates that other data show up-regulation of this cotton gene in transition-stage fiber (Wu and Liu 2005). Arabidopsis gene names and At Gene IDs are shown on the y-axis, and names for corresponding Arabidopsis mutants are italicized beside the histogram bars. The designations 1° and/or 2° indicate whether the genes have been associated with primary and/or secondary wall synthesis in Arabidopsis (and/or other species) through bioinformatics, analysis of mutant phenotypes, or both. Regarding gene names, two Arabidopsis homologs are shown for tubulin and sucrose synthase because each of these pairs represents closely related isoforms with equal homology to the cotton cDNA spotted on the array

foundational event for the diversification of land plants (Bowman et al. 2007). Figure 6 shows a fusion of Arabidopsis gene names and mutant information with gene expression data (histogram bars) in isolated fiber for the *cotton homolog* of the Arabidopsis gene (see legend for further details). The homology relationships, and possible orthology, were confirmed by reciprocal BLAST of the longest cotton unigene vs. the Arabidopsis proteome and the identified Arabidopsis protein vs. the cotton comprehensive gene assembly.

Included in Fig. 6 are many genes known to be required for cell wall synthesis, many identified through mutant analysis and a few through bioinformatics alone. Mutant series that appear more than once in Fig. 6 are: (a) *irx*, or irregular/collapsing xylem resulting from weak secondary walls in water-conducting cells; (b) *fra*, derived from the fragile inflorescence stems of mutants with thin secondary walls in interfascicular fibers; and (c) *rsw*, derived from a root radial swelling phenotype. The *irx* and *fra* phenotypes have been particularly useful in revealing genes required for secondary wall deposition. Note that the *ixr1* mutant name is different from *irx*, denoting resistance to the

cellulose-synthesis inhibitor isoxaben. Three bioinformatics studies also formed the basis for Fig. 6. One comparative transcriptome analysis focused on identifying a “core xylem gene set” already highlighted xylem-expressed genes with homologs in cotton fiber (Ko et al. 2006). Some of these genes that had G.h.fbr-sw homologs spotted on the array and about which we have some functional information are included in Fig. 6. Two analyses of genes co-expressed in *Arabidopsis* with secondary wall CESA genes identified a total of 52 unique sequences (Brown et al. 2005; Persson et al. 2005). These studies were recently synthesized to show a set of 19 commonly identified secondary-wall-associated genes, along with any collapsing xylem (*irx*) phenotypes observed and/or cellulose depletion documented by other means (Table 1 in Joshi and Mansfield 2007). ESTs representing cotton homologs of 13 of these 19 genes were sequenced from 20 DPA cotton fiber and included on the microarray (see gene names with asterisks in Fig. 6). (An EST from a 14th cotton homolog, for the unknown protein At1g09610, was in the G.h.fbr-sw sequences but did not amplify for the spotted cDNA array. The other 5 genes among the 19 genes dually found to be co-regulated with *Arabidopsis* secondary wall CESA genes were not the closest homolog of a known cotton fiber gene (data not shown.)

7 Function of Genes Related to Secondary Wall Deposition

Twenty-one of the genes in Fig. 6 were associated with secondary wall deposition in xylem (see 2° notation), and 15 of the cotton fiber homologs showed up-regulation at the onset of secondary wall deposition on the spotted cDNA microarray. The functional context of these genes is briefly summarized here. For findings from species other than cotton, often only one recent publication is cited to provide an entry point into the relevant literature. Information from *Arabidopsis* is emphasized, but many parallels also exist between gene expression for cotton fiber secondary wall deposition and formation of woody stem tissue in trees and grain stalks (e.g. see Hertzberg et al. 2001). Acknowledging the ultimate need for confirmation of each gene expression pattern by qPCR, we will omit the word “putatively” in reference to gene expression patterns discussed further here. Also, some genes may have failed to show significant up-regulation on the spotted cDNA microarray, whereas changes would be detected by a gene-specific oligomer arrays or qPCR. Finally, increased gene expression does not always correlate with translation or increased protein activity, but other evidence supports the similarity of secondary wall deposition in cotton fiber and xylem. In addition to similar results for cotton (Zhang et al. 2004b), when the promoter of secondary-wall-expressed *GhCESA4* was used in a reporter gene construct in transgenic *Arabidopsis*, GUS activity was observed in a variety of *Arabidopsis* cells with thickened walls, including the xylem (Kim and Triplett 2001).

7.1 *Specific Isoforms Within Families Are Up-regulated*

First, we describe three cases illustrating isoform-specific involvement within one protein family in secondary wall deposition in cotton fiber, as also occurs in xylem of other species. The *CESA* genes encode proteins included in the CSC that mediates the biogenesis of crystalline cellulose nanofibrils. These Family 2 processive glycosyltransferases can add UDP-glc to continue the elongation of a β -1,4-glucan chain. Arabidopsis has 10 related genes (*AtCESA1-10*), and three genes have non-redundant roles for primary wall synthesis (*AtCESA1*, -3, -6) or for secondary wall synthesis (*AtCESA4*, -7, -8). In addition, *AtCESA2* is required for normal cellulose synthesis in primary walls of some aerial Arabidopsis organs (Chu et al. 2007). Now, CESAs from many plant species have been shown to group into clades of “primary wall type” and “secondary wall type” proteins. Increasing evidence supports the existence of a multimeric CSC in which particular CESA isoforms occupy non-interchangeable sites (Taylor et al. 2003; Wang et al. 2006). In the fiber microarray data, the cotton homologs of *AtCESA4*, -7, -8 were strongly up-regulated at the transition to secondary wall deposition on 20 DPA. The expression of *GhCESA1* (Accession U58283), the cotton homolog of *AtCESA8*, was shown previously to be strongly up-regulated at the onset of fiber secondary wall synthesis (Pear et al. 1996) and the encoded protein also appeared at that time (Peng et al. 2002). In contrast, the cotton homologs of at least two Arabidopsis primary-wall-related *CESAs*, *AtCESA1* and *AtCESA3*, were still being transcribed at 20 DPA, but they were not up-regulated since fiber elongation via primary wall synthesis was slowing or stopped. Supporting these results, a cotton gene with 98% nucleotide identity to the Gh.fbr.sw EST that was homologous to *AtCESA3* was previously sequenced from 24 DPA fiber (*GhCESA3*, Accession AF150630). However, preliminary evidence suggested that its expression peaked during fiber elongation (Laosinchai 2002). Probably arising from cotton allotetraploidy and/or emphasis on cellulose synthesis in cotton fiber, *GhCESA4* (AF413210), a homolog of *GhCESA1* with 96% nucleotide identity to it, was reported to have secondary-wall specific expression in fiber (Kim et al. 2002). However, another *GhCESA1*-type sequence (AF139448, called at first *CelA2*) shows evidence of pseudogenization that originated in the ancestral diploid A-genome and persisted in allotetraploid *G. hirsutum* (Cronn et al. 1999). In analogy with other crop plants (reviewed in Farrokhi et al. 2006), *Gossypium* is likely to contain numerous distinct CesA genes, with changes with possible functional consequences having occurred within species.

A second case of isoform-specific up-regulation for secondary wall deposition occurs within the *COBRA* gene family. COBRA proteins are of unknown function, but they localize to the external face of the plasma membrane and may function coordinately with microtubules as well as binding to a cell wall polysaccharide. Loss of COBRA was associated with abnormal anisotropic cell expansion, cellulose deficiency, and poorly oriented cellulose fibrils in

Arabidopsis roots. The failure of its cotton fiber homolog to up-regulate for secondary wall deposition can be explained as for *AtCESA1* and *AtCESA3*. In contrast, the cotton homolog of Arabidopsis *COBRA-LIKE4* was up-regulated for fiber secondary wall deposition, which is consistent with orthologous genes being required to achieve the normal degree of xylem secondary wall deposition in rice, corn, and Arabidopsis (Brady et al. 2007). A third case occurs within a family of chitinase-like genes. Mutation in *POM1/AtCTL1* was associated with defects in root development, incomplete primary walls, and pleiotropic stress phenotypes (Zhong et al. 2002) as well as cellulose deficiency in Arabidopsis (Piling and Hofte 2003). Subsequently, *AtCTL2* was revealed by bioinformatics to be associated with expression of Arabidopsis secondary-wall-*CESA* genes, and this gene is the closest match to cotton genes with secondary-wall-enhanced expression as shown by Northern blotting and promoter:GUS assay (Zhang et al. 2004b), as well as by microarray and qPCR analysis (Fig. 5, 6). The expression patterns of two cotton *CTL*-type genes on the array correspond to the primary- or secondary-wall roles of the two (apparent) Arabidopsis orthologs.

7.2 *Transcriptional Control*

The discussion now continues approximately in order of how up-regulated genes are likely to function in the transition to cotton fiber secondary wall deposition. Clearly there are transcription factors that mediate the large-scale gene transcriptional changes that occur during this developmental shift. One of these is homologous to the Arabidopsis homeodomain transcription factor, *KNAT7* (knotted1-like7). Typically, *KNAT* transcription factors regulate development via networks of protein interactions (Truernit et al. 2006). In Arabidopsis xylem, this transcription factor acts downstream of NAC domain transcription factors that are master regulators of the whole process of xylem secondary wall deposition, including cellulose, lignin, and xylan synthesis (Zhong et al. 2007). In cotton fiber, lignin synthesis does not occur, so it is tempting to speculate that *KNAT7* homologs have a specific role in stimulating secondary wall cellulose biogenesis. However, we cannot disallow a possible role in xylan synthesis at this time (see below).

7.3 *Microtubule-related Genes*

Among the first visible changes at the transition to secondary wall deposition in cotton fiber are changes in the angle of microtubules and cellulose fibrils. As secondary wall synthesis progresses, microtubules become more numerous, longer, closer to the plasma membrane, and oriented in a more steeply pitched helix that is mirrored by cellulose nanofibril orientation (Seagull 1993). The

expression of tubulin was also up-regulated for secondary wall synthesis (Fig. 6), which correlates with increasing microtubule density, and, possibly, functional differences between tubulin isoforms. Also related to microtubule function, the expression is up-regulated of the cotton homolog of the p60 subunit of *KATANIN*, which encodes a widely-expressed and broadly-important AAA-ATPase with microtubule-severing capacity. Among diverse phenotypes in plants, some of its mutant alleles decreased the quantity and organization of cellulose in secondary walls of Arabidopsis interfascicular fibers (Burk and Ye 2002). Microtubule severing likely aids in reconfiguring the microtubule array to support the deposition of more cellulose in a different orientation in secondary walls compared to primary walls. Probably for similar reasons, a cotton homolog of *SPIRAL1-LIKE5* was up-regulated at 20 DPA. *SPIRAL1* (*SPR1*) has unknown biochemical function, but it localizes to microtubules and mediates directional growth of rapidly expanding cells in diverse plant organs. All *SPIRAL1-LIKE* genes are at least partly functionally redundant with *SPIRAL1* in Arabidopsis, but double mutants lead to randomly oriented cortical microtubule arrays (Nakajima et al. 2006). Clearly, orderly microtubule arrays are important for cotton secondary wall deposition.

7.4 Synthesis of Non-Cellulosic Polysaccharides

The cotton homolog of At2g28110 (*fra8*, *irx7*), in glycosyltransferase family 47, was up-regulated on the array. Although mutations caused both cellulose and xylan deficiency in Arabidopsis xylem, this enzyme participates in the synthesis of a special 5-sugar terminator on the reducing end of glucuronoxylan (Pena et al. 2007). In addition, the cotton homolog of *UXS6*, a UDP-D-glucuronate decarboxylase yielding UDP-D-xylose, was up-regulated on the array. This nucleotide sugar can provide xylose for biosynthesis of xyloglucan and arabinoxylan, as well as undergo epimerization to UDP-L-arabinose for arabinoxylan and pectin biosynthesis. Previously some data implicated an increase in xylose during cotton fiber secondary wall formation (reviewed in Ryser 1985), but, at most, only minor amounts of any typical cell wall matrix molecules exist in this thick wall. However, the transitional “winding” layer, which is similar to the S1 layer in a wood fiber, may contain molecules not yet characterized in detail in cotton. The possible existence of a xylan-type molecule requires explanation of why the cotton homolog of At2g37090 (*irx9*) was not shown to be significantly up-regulated on the array. This gene encodes a Family 43 beta-glucuronyltransferase involved in glucuronoxylan backbone elongation (Pena et al. 2007). Microarray results are never completely accurate, and it is also possible that 20 DPA was too late to capture the up-regulation of the cotton homolog of At2g37090 as shown by other results. The G.h.fbr-sw cDNA spotted on the array is contained within *GhGlcAT1* (*G. hirsutum* glucuronic acid transferase 1, AY346330), which was shown by Northern blotting to have a

sharp peak of expression in 15 DPA cotton fiber as well as being expressed in stems that contain xylem (Wu and Liu 2005). The histogram bar for the cotton At2g37090 homolog was striped in Fig. 6 to indicate these prior results. Therefore, cotton fiber does express genes closely related to those required for the synthesis of glucuronoxylan in xylem cells. In this case, analysis of gene expression drives the need to look in detail for cellular mechanisms not yet fully characterized in cotton fiber development.

7.5 Supply of Carbon to Cellulose Synthetic Enzymes

Sink strength strongly increases for secondary wall deposition, and the expression of sucrose synthase (*SUS* (called previously *SuSy*; E.C. 2.4.1.13; $\text{suc} + \text{UDP} \longleftrightarrow \text{UDP-glc} + \text{fru}$) increases to support the supply of more UDP-glc to CESA enzymes as shown previously by Northern blotting (Amor et al. 1995) and also in Fig. 5, 6. Increased *SUS* gene expression is correlated with increased *SUS* protein near the cortical microtubules and plasma membrane (the site of CSCs) as cotton fiber secondary wall deposition commences, as well as in the exoplasmic space where callose (β -1,3-glucan) is deposited and cellulose fibrils crystallize (Salnikov et al. 2003). Recently, maize *SUS1* was shown to have intrinsic membrane-binding domains, with membrane association promoted by high *SUS1* protein and sucrose concentration and lower pH (Hardin et al. 2006). In *Arabidopsis*, genetic analyses provided no evidence that a particular *SUS* isoform was specifically associated with cellulose synthesis; even a *sus1sus4* double mutant grew normally except when flooded (Bieniawska et al. 2007). However, regulation of the intracellular behavior of *SUS* transcripts or protein may modulate cellulose synthesis under normal conditions in all plants (Haigler et al. 2001). Furthermore, crop plants that have been domesticated to develop large storage organs have shown detrimental effects of down-regulation of particular *SUS* isoforms (Bieniawska et al. 2007). The proven role of *SUS* in cotton fiber initiation and elongation based on anti-sense suppression of cotton *SUS* (U73588) has been reviewed recently, including preliminary evidence that transgenic suppression of *SUS* reduces fiber secondary wall cellulose content (Ruan 2007). It will be interesting to explore the functional significance of the putative sucrose transporter, *GhSUC*, that is up-regulated along with secondary wall cellulose synthesis (Fig. 5), especially because another fiber sucrose transporter (*GhSUT1* with unspecified sequence, Ruan 2007) has maximal expression during elongation. Furthermore, current models suggest that there is less need for sucrose to be transported across membranes when plasmodesmata are open at the fiber foot during secondary wall deposition (Ruan 2007).

Sucrose degradation by *SUS* releases only half the carbon as UDP-glc to be used immediately by CESA enzymes. To recycle the coordinately-released fructose (after its phosphorylation) toward additional sucrose to support the extensive cellulose synthesis during cotton fiber secondary wall deposition,

sucrose phosphate synthase (SPS; E.C. 2.4.1.14; $\text{fru-6-P} + \text{UDP-glc} \rightarrow \text{suc-P} + \text{UDP}$) may play a role. Hindrance in the intra-fiber synthesis of sucrose is correlated with the limitation of cellulose synthesis rate by cool temperatures, which in turn can result in fiber immaturity in the field (Martin and Haigler 2004). Transgenic cotton with SPS activity elevated in leaves and, to a lesser extent, in fiber showed increased fiber secondary wall deposition in growth chamber experiments (Haigler et al. 2007). Nonetheless, one native cotton *SPS* gene assayed by qPCR (Fig. 5) showed down-regulation for secondary wall synthesis. Cotton is likely to contain several SPS isoforms (Lutfiyya et al. 2006), and additional data are needed to clarify whether or not one of them has a particular relationship to secondary wall cellulose synthesis.

In addition, it is important to remember that metabolism is often controlled post-transcriptionally. Even though the UDP-glc pyrophosphorylase gene assayed here did not show increased expression for secondary wall deposition (Fig. 5), the size of the UDP-glc pool does increase at this time. There is likely metabolic rather than transcriptional control of UDPG_P enzyme activity, and SPS activity is also highly regulated by phosphorylation and interaction with co-factors and other proteins (Haigler et al. 2001).

7.6 Genes of Unknown Function

The cotton homolog of the gene for KORRIGAN is also up-regulated on the array, which is consistent with a large increase in KORRIGAN protein at the onset of cotton fiber secondary wall deposition (Peng et al. 2002). In Arabidopsis, KORRIGAN, which has cellulase or 1,4- β -D-glucanase activity *in vitro* but an unknown substrate *in vivo*, is required for both primary and secondary wall synthesis. Severe phenotypes arise from its functional loss. KORRIGAN cycles between the endomembrane system and the plasma membrane in a microtubule-dependent manner, but available evidence suggests that it is not part of the CSC that contains CESA proteins (Robert et al. 2005; Szyjanowicz et al. 2004). Speculations about the role of KORRIGAN include cleaving a primer for cellulose synthesis, editing glucan chains prior to cellulose crystallization, cleaving glucan chains to release tensional stress of cellulose crystallization, or releasing CSCs from the cellulose fibril.

Other genes up-regulated on the array have even less known about their function. (1) The cotton homolog of At4g27435 (AAR07956; Fb34) was associated with an informative EST-SSR (expressed sequence tag simple sequence repeat) in a mapping population derived from a cross of *G. hirsutum* and *G. barbadense*, with the latter producing the highest quality cotton fiber (Han et al. 2004). The poplar homolog of this gene is up-regulated during wood secondary wall formation, and one Arabidopsis knock-out mutant showed about 20% less secondary xylem production under special growth conditions that induce that process in Arabidopsis (Ubeda-Tomas et al. 2007). (2) The

up-regulated fasciclin-like protein FLA11 (*irx13*) belongs to a group of arabinogalactan proteins (AGPs); these secreted proteoglycans contain only 1–10% protein. Homologous genes to Arabidopsis FLA11 and FLA12 are strongly expressed in cellulose-rich tension wood (Lafarguette et al. 2004), and a gene with closest homology to Arabidopsis FLA12 was sequenced from 10 DPA cotton fiber and shown not to be present in ovules of wild-type cotton or a fiberless mutant (GhAGP1, AAO92753; Ji et al. 2003). The Gh.fbr.sw EST spotted on the array is nearly identical to *GhAGP1*, and further tests by qPCR would be needed to determine if true up-regulation of this particular AGP occurs for cotton fiber secondary wall deposition since the increase shown for the FLA12 cotton homolog at 20 DPA by the microarray was non-significant. Based on the presence of their putative cell adhesion domains (fasciclin domains) and immunolocalization in tension wood at the interface between the cell wall and plasma membrane [to which AGPs may be attached through glycosylphosphatidylinositol (GPI) anchors], it was proposed that the AGPs may act as adhesive proteins or participate in a signaling networks based on N-acetyl-glucosamine (Lafarguette et al. 2004).

7.7 Cell-wall Related Genes Not Shown to be Up-regulated

If not already mentioned, below is a brief summary of the functional context in other types of genes that were not up-regulated on the array. Additional experiments (including checking the gene expression pattern by qPCR) would be required to test whether or not these or related proteins have similar roles in cotton fiber. (1) In Arabidopsis, RHD3, which has GTP-binding motifs, is thought to be involved in endomembrane trafficking of molecules required for cell wall synthesis, thereby helping to control anisotropic cell expansion as well as secondary wall thickening (Hu et al. 2003; Yuen et al. 2005). (2) Down-regulation of *PGSIP1*, a glycosyltransferase family 8 gene, in Arabidopsis is associated with starch deficiency, and the protein likely primes starch synthesis. Starch is not present after early elongation in cotton fiber (Ryser 1985), but the expression pattern of the poplar homolog suggests a possible role in priming synthesis of another polysaccharide during wood secondary wall formation (Aspeborg et al. 2005). (3) The serine-rich protein, KOBITO, in Arabidopsis has a role in cell elongation, cellulose synthesis, vascular differentiation, and root meristem maintenance. However, multiple down-stream phenotypes arose from alteration of an ABA- and Glc-responsive signaling network (Lertpiriyapong and Sung 2003; Brocard-Gifford et al. 2004), and the same regulatory network might not operate in cotton fiber. (4) For the multi-copper oxidase, *LACCASE4*, down-regulation of a close poplar homolog resulted in increased soluble phenolics in xylem parenchyma cells and misshapen secondary-walled conducting cells, which also had secondary walls that were poorly adhering internally or to the adjacent primary wall (Ranocha et al. 2002). These authors

inferred that polymerization of non-lignin phenolics was important to maintain xylem secondary wall integrity. (5) For the glycosyltransferase family 47 gene (*At1g27440*), the poplar homolog is related to cell wall cross-linking in wood. Based on analysis of a tobacco homolog, it may be that this gene modifies pectin type II rhamnogalacturonan by addition of glucuronic acid, which in turn can affect borate cross-linking (Ubeda-Tomas et al. 2007). Neither phenolic or borate cross-linking are known to be important in secondary wall cotton fibers.

The secondary-wall-biased G.h.fbr-sw sequence set did not include cotton homologs of several genes that support primary wall cellulose synthesis (and probably secondary wall synthesis) at a fundamental level in *Arabidopsis* including: (a) *CYT1* and *KNF* that, along with *RSW3*, encode proteins that carry out N-glycan processing in the endoplasmic reticulum; (b) *FK*, *HYD1*, and *SMT1/CPH* that encode proteins involved in sterol biosynthesis (Schrick et al. 2004); and (c) *RSW10*, encoding a ribose 5-phosphate isomerase that is likely generates UDP-glc (see review in Somerville 2006 and Howles et al. 2006). Probably their absence from these experiments denotes their constitutive expression in cotton fiber.

8 Conclusions

We have illustrated the power of cotton fiber gene expression to reveal genes that are required across species for secondary wall deposition, particularly high-rate cellulose synthesis. We showed that the highly developed genomic resources for *Arabidopsis* can aid functional understanding of cotton genes. The similarities of cotton fiber secondary wall gene expression with xylem reveal an opportunity to analyze the “pure system” of cotton fiber for as yet unknown participants in cellulose synthesis. In addition to up-regulated genes, there are surely constitutively expressed genes that also have important roles. Further research on secondary wall deposition in cotton fiber is likely to be beneficial to the improvement of numerous cellulose-rich fiber crops.

In addition, one of the most interesting challenges is to understand what makes cotton fiber different, e.g. why does it not synthesize large amounts of matrix molecules in its secondary wall? What is the role of proteins encoded by cotton-specific and fiber-specific genes? Are there any unique aspects of cellulose synthesis in the fiber? How are the developmental transitions in the fiber controlled and interfaced with environmental signals? How is cellulose synthesis down-regulated under stress?

Similar analyses to those reported here using cDNA arrays have recently been reported for *G. hirsutum* (Guo et al. 2007) and *G. barbadense* (Tu et al. 2007), and soon we will see the application of more comprehensive gene-specific oligomer arrays to understand changes in gene expression at the onset of secondary wall deposition. All such research quickly results in the need to assign gene function through experimentation, a process that is much slower than

demonstrating changes in gene expression. Achieving functional understanding will require more integration of gene expression studies with, for example, analysis of protein-protein interactions, enzymatic activity, metabolite levels and flux, and cell structure via advanced imaging techniques. More rapid means of testing gene function in cotton, e.g. by more efficient and faster production of stably transformed cotton or by alternative means such as virus induced gene silencing, would certainly be useful.

Given that the cotton genome sequence is not yet available, molecular means of cotton improvement must rely on the existing large body of cotton ESTs and ability to analyze the expression and significance of genes they represent. Such information can help bridge cotton and model plants, including Arabidopsis and Poplar, for comparative genome analysis. It can also be directly used for genetic improvement of fiber quality and yield traits. For example, if fiber ESTs can be associated with fiber traits by QTL mapping, fine mapping of the QTL region can lead to development of molecular markers for marker-assisted breeding.

Acknowledgments Cotton Incorporated, Cary, NC supported the microarray and bioinformatics analyses, with additional support from North Carolina State University. Special thanks go to Curtis Wilkerson and Jeff Landgraf, Research Technology Support Facility/Bioinformatics at Michigan State University, for carrying out the microarray analysis. NSF Plant Genome Research Program grants, #DBI-0211797, R98RA1829, and #DBI-0110173, supported making and sequencing the G.h.fbr-sw SSH library.

References

- Adams, K.L. and Wendel, J.F. (2004) Exploring the genomic mysteries of polyploidy in cotton. *Biol. J. Linn. Soc.* 82: 573–581.
- Amor Y., Haigler, C.H., Johnson, S., Wainscott, M. and Delmer, D.P. (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proc. Natl. Acad. Sci. USA* 92: 9353–9357.
- Arpat, A., Waugh, M.P., Sullivan, J., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R. and Wilkins, T. (2004) Functional genomics of cell elongation in developing cotton fibers. *Plant Mol. Biol.* 54: 911–929.
- Arthur, J.C. (1990) Cotton. In: J.I. Kroschwitz (Ed.), *Polymers: Fibers and Textiles, a Compendium*, John Wiley and Sons, New York, pp. 118–141.
- Aspeborg, H., Schrader, J., Coutinho, P.M., Stam, M., Kallas, A., Djerbi, S., Nilsson, P., Denman, S., Amini, B., Sterky, F., Master, E., Sandberg, G., Mellerowicz, E., Sundberg, G., Henrissat, B. and Teeri, T.T. (2005) Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiol.* 137: 983–997.
- Baskin, T.I. (2001) On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. *Protoplasma* 215: 150–171.
- Benedict, C.R., Kohel, R.J. and Jividen, G.M. (1994) Crystalline cellulose and cotton fiber strength. *Crop Sci.* 34: 147–151.
- Benedict, C.R., Kohel, R.J. and Lewis, H.L. (1999) Cotton fiber quality, In: C.W. Smith, J.T. Cothren (Eds.), *Cotton: Origin, History, Technology, and Production*, John Wiley & Sons, New York, pp. 269–288.

- Bieniawska, Z., Paul Barratt, D.H., Garlick, A.P., Thole, V., Kruger, N.J., Martin, C., Zrenner, R. and Smith, A.M. (2007) Analysis of the sucrose synthase gene family in *Arabidopsis*. *Plant J.* 49: 810–828.
- Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* 6:791–806.
- Bowman, J.L., Floyd, S.K. and Sakakibara, K. (2007) Green genes—comparative genomics of the green branch of life. *Cell* 129: 229–234.
- Bradow, J.M. and Davidonis, G.W. (2000) Review: Quantitation of fiber quality and the cotton production-processing interface: A physiologist's perspective. *J. Cotton Sci.* 4: 34–64.
- Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A. and Benfey, P.N. (2007) Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant Physiol.* 143:172–87.
- Brocard-Gifford, I., Lynch, T. J., Garcia, E. M., Malhotra, B. and Finklestein, R. R. (2004) The *Arabidopsis thaliana* abscisic acid insensitive8 locus encodes a novel protein mediating abscisic acid and sugar responses essential for growth. *Plant Cell* 16: 406–421.
- Brown, D.M., Zeef, L. A.-H., Ellis, J., Goodacre, R. and Turner, S.R. (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17: 2281–2295.
- Brown, R.M. Jr. and Saxena, I.M. (Eds.) (2007) *Cellulose: Molecular and Structural Biology*. Springer, Dordrecht, 379 pp.
- Burk, D.H. and Ye, Z.-H. (2002) Alteration of oriented deposition of cellulose microfibrils by mutation of a katanin-like microtubule severing protein. *Plant Cell* 14: 2145–2160.
- Busch, W. and Lohmann, J.U. (2007) Profiling a plant: expression analysis in *Arabidopsis*. *Curr. Opin. Plant Biol.* 10: 136–141.
- Carpita, N.C. and Delmer, D.P. (1981) Concentration and metabolic turnover of UDP-glucose in developing cotton fibers. *J. Biol. Chem.* 256: 308–315.
- Chanzy, H., Imada, K. and Vuong, R. (1978) Electron diffraction from the primary wall of cotton fibers. *Protoplasma* 94: 299–306.
- Chu, A., Chen, H., Zhang, Y., Zhang, Z., Zheng, N., Yin, B., Yan, H., Zhu, L., Zhao, X., Yuan, M., Zhang, X. and Xie, Q. (2007) Knockout of the *AtCESA2* gene affects microtubule orientation and causes abnormal cell expansion in *Arabidopsis*. *Plant Physiol.* 143: 213–224.
- Cronn, R.C., Small, R.L. and Wendel, J.F. (1999) Duplicated genes evolve independently after polyploidy formation in cotton. *Proc. Natl. Acad. Sci. USA* 96: 14406–14411.
- Delmer, D.P. (1999) Cellulose biosynthesis in developing cotton fibers. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 85–112.
- Diatchenko, L., Lau, Y.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025–6030.
- Diotallevi, F. and Mulder, B. (2007) The cellulose synthase complex: a polymerization driven supramolecular motor. *Biophys. J.* 92: 2666–2673.
- Endrizzi, J.E., Turcotte, E.L. and Kohel, R.J. (1985) Genetics, cytology, and evolution of *Gossypium*. *Adv. Gen.* 23: 271–354.
- Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A. and Fincher, G.B. (2006) Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to identification of key genes. *Plant Biotech. J.* 4: 145–167.
- Franz, G. (1969) Soluble nucleotides in growing cotton hair. *Phytochem.* 8: 737–741.
- Fryxell, P.A. (1979) *The Natural History of the Cotton Tribe (Malvaceae, Tribe Gossypieae)*. Texas A&M University Press, College Station, 245 pp.

- Gipson, J.R. (1986) Temperature effects on growth, development, and fiber properties. In: J.R. Mauney and J. McD. Stewart (Eds.), *Cotton Physiology*, The Cotton Foundation, Memphis, pp. 47–56.
- Guo, J.-Y., Wang, L.-J., Chen, S.-P., Hu, W.-L. and Chen, X.-Y. (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary wall synthesis. *Cell Research* 2007: 1–13.
- Gutierrez, R.A., Green, P.J., Keegstra, K. and Ohlrogge, J.B. (2004) Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms. *Genome Biology* 5:R53 (<http://genomebiology.com/2004/5/8/R53>).
- Haigler, C.H. (1985) The functions and biogenesis of native cellulose. In: S.H. Zeronian and T.P. Nevell TP (Eds.), *Cellulose Chemistry and its Applications*, Ellis Horwood, Chichester, pp 30–83.
- Haigler, C.H. (1991) The relationship between polymerization and crystallization in cellulose biogenesis. In: C.H. Haigler and P. Weimer (Eds.), *Biosynthesis and Biodegradation of Cellulose*, Marcel Dekker, New York, pp 99–124.
- Haigler, C.H., Ivanova-Datcheva, M., Hogan, P.S. Salnikov, V.V., Hwang, S., Martin, L.K. and Delmer, D.P. (2001) Carbon partitioning to cellulose synthesis. *Plant Mol. Biol.* 47: 29–51.
- Haigler, C.H., Zhang, D. and Wilkerson, C.G. (2005) Biotechnological improvement of cotton fibre maturity. *Physiol. Plant.* 124: 285–294.
- Haigler, C.H. (2006) Establishing the cellular and biophysical context of cellulose synthesis. In: T. Hayashi (Ed.), *The Science and Lore of the Plant Cell Wall: Biosynthesis, Structure and Function*, BrownWalker Press: Boca Raton, pp. 97–105.
- Haigler, C.H. (2007) Substrate supply for cellulose synthesis and its stress sensitivity in the cotton fiber. In: R.M. Brown Jr, and I. Saxena (Eds.), *Cellulose: Molecular and Structural Biology*, Springer: New York, pp. 145–166.
- Haigler, C.H., Singh, B., Zhang, D., Hwang, S., Wu, C., Cai, W.X., Hozain, M., Kang, W., Kiedaisch, B., Strauss, R.E., Hequet, E.F., Wyatt, B.G., Jviden, G.M. and Holaday, A.S. (2007) Transgenic cotton over-producing spinach sucrose phosphate synthase showed enhanced leaf sucrose synthesis and improved fiber quality under controlled environmental conditions. *Plant Mol. Biol.* 63: 815–832.
- Han, Z.G., Guo, W.Z., Song, X.L. and Zhang, T.Z. (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Mol. Gen. Genom.* 272: 308–327.
- Hardin, S.C., Duncan, K.A. and Huber, S.C. (2006) Determination of structural requirements and probable regulatory effectors for membrane association of maize sucrose synthase. *Plant Physiol.* 141: 1106–1119.
- Heinze, T. (1998) New ionic polymers by cellulose functionalization. *Macromol. Chem. Phys.* 1999: 2341–2364.
- Hertzberg, M., Aspeborg, H., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlén, M., Teeri, T.T., Lundeberg, J., Sundberg, B., Nilsson, and P. Sandberg, G. (2001) A transcriptional roadmap to wood formation. *Proc. Natl. Acad. Sci. USA* 98: 14732–14737.
- Hindeleh, A.M., Johnson, D.J. and Monatgue, P.E. (1980) Computational methods for profile resolution and crystallite size evaluation in fibrous polymers. In: A.D. French and K.H. Gardener (Eds.), *Fiber Diffraction Methods: ACS Symposia no. 141*, p. 149–182.
- Howles, P.A., Birch, R.J., Collings, D.A., Gebbie, L.K., Hurley, U.A., Hocart, C.H., Arioli, T. and Williamson, R.E. (2006) A mutation in an *Arabidopsis* ribose 5-phosphate isomerase reduces cellulose synthesis and is rescued by exogenous uridine. *Plant J.* 48: 606–618.
- Hsieh, Y.-L. (1999) Structural development of cotton fibers and linkages to fiber quality. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 137–166.

- Hsieh, Y.-L., Hu, X.P. and Nguyen, A. (1997) Strength and crystalline structure of developing Acala cotton. *Text. Res. J.* 67: 529–536.
- Hu, X.P. and Hsieh, Y.-L. (1996) Crystalline structure of developing cotton fibers. *J. Polym. Sci.: Part B: Polym. Phys.* 34: 1451–1459.
- Hu, H.Y., Zhong, R., Morrison, W. and Ye, Z. (2003) The Arabidopsis RHD3 gene is required for cell wall biosynthesis and actin organization. *Planta* 217: 912–921.
- Hutchinson, J.B., Stephens, S.G. and Dodds, K.S. (1945) The seed hairs of *Gossypium*. *Ann. Bot.* IX (36): 360–368.
- Jacob-Wilk, D., Kurek, I., Hogan, P. and Delmer, D.P. (2006) The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* 103: 12191–12196.
- Ji, S.J., Lu, Y.C., Feng, J.X., Wei, G., Li, J., Shi, Y.H., Fu, Q., Liu, D., Luo, J.C. and Zhu, Y.X. (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res.* 31: 2534–2543.
- Jiang, C., Wright, R.J., El-Zik, K.M. and Paterson, A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium*. *Proc. Natl. Acad. Sci. USA* 95(8): 4419–4424.
- Joshi, C.P. and Mansfield, S.D. (2007) The cellulose paradox—simple molecule, complex biosynthesis. *Curr. Opin. Plant Biol.* 10: 220–226.
- Ko, J.-H., Beers, E.P. and Han, K.H. (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 276: 517–531.
- Kim, H.J. and Triplett, B.A. (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127: 1361–1366.
- Kim, H.J., Williams, M.Y. and Triplett, B.A. (2002) A novel expression assay system for fiber-specific promoters in developing cotton fibers. *Plant Mol. Biol. Rep.* 20: 7–18.
- Kurek, I., Kawogoe, Y., Jacob-Wilk, D., Doblin, M. and Delmer, D. (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domain. *Proc. Natl. Acad. Sci.* 99: 11109–11114.
- Lafarguette, F., Leple, J.-C., Dejardin, A., Laurans, F., Costa, G., Lesage-Descauses, M.-C. and Pilate, G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 164: 107–121.
- Laosinchai, W. (2002) Molecular and biochemical studies of cellulose and callose synthase. Ph.D. Dissertation, The University of Texas at Austin, 207 pp.
- Lertpiriyapong, K. and Sung, Z. R. (2003) The elongation defective1 mutant of *Arabidopsis* is impaired in the gene encoding a serine-rich secreted protein. *Plant Mol. Biol.* 53: 581–595.
- Lewin, M. and Pearce, E.M. (Eds.) (1998) *Handbook of Fiber Chemistry*, 2nd edn, Marcel Dekker: NY, 724 pp.
- Lutfiyya, L.L., Xu, N., D'Ordine, R.L., Morrell, J.A., Miller, P.W. and Duff, S.M.G. (2006) Phylogenetic and expression analysis of sucrose phosphate synthase isozymes in plants. *J. Plant Physiol.* Doi:10.1016/j.jpiph.2006.04.014.
- Martin, L.K. and Haigler, C.H. (2004) Cool temperature hinders flux from glucose to sucrose during cellulose synthesis in secondary wall stage cotton fibers. *Cellulose* 11:339–349.
- Meinert, M.C. and Delmer, D.P. (1977) Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiol.* 59: 1088–1097.
- Nakajima, K., Kawamura, T. and Hashimoto, T. (2006) Role of the *SPIRAL1* gene family in anisotropic growth of *Arabidopsis*. *Plant Cell Physiol.* 47: 513–522.
- Niklas, K.J. (1992) *Plant Biomechanics, An Engineering Approach to Plant Form and Function*, Univ Chicago Press, Chicago, 607 pp.
- Paradez, A.R., Someville, C.R. and Erhardt, D.W. (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* 312: 1491–1495.

- Pear, J., Kawagoe, Y., Schreckengost, W., Delmer, D.P. and Stalker, D. (1996) Higher plants contain homologs of the CelA genes that encode the catalytic subunit of the bacterial cellulose synthases. *Proc. Natl. Acad. Sci. USA* 93: 12637–12642.
- Pena, M.J., Zhong, R., Zhou, G.-K., Richardson, E.A., O'Neill, M.A., Darvill, A.G., York, W.S. and Ye, Z.-H. (2007) *Arabidopsis irregular xylem8* and *irregular xylem 9*: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19: 549–563.
- Peng, L., Kawagoe, Y., Hogan, P. and Delmer, D. (2002) Sitosterol- β -glucoside as primer for cellulose synthesis in plants. *Science* 295: 147–150.
- Persson S., Wei, H., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. USA* 102: 8633–8638.
- Pilate, G., Dejardin, A., Laurans, F. and Leple, J.-C. (2004) Tension wood as a model for functional genomics of wood formation. *New Phytol.* 164: 63–72.
- Piling, E. and Hofte, H. (2003) Feedback from the wall. *Curr. Opin. Plant Biol.* 6: 611–616.
- Potikha, T.S., Collins, C.C., Johnson, D.I., Delmer, D.P. and Levine, A. (1999) The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. *Plant Physiol.* 119: 849–858.
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, URL <http://www.R-project.org>.
- Ranocha, P., Chabannes, M., Chamayou, S., Danoun, S., Jauneau, A., Boudet, A.-M. and Goffner, D. (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiol.* 129: 145–155.
- Rebenfield, L. (1990) Fibers. In: J.I. Kroschwitz (Ed.), *Polymers: Fibers and Textiles, a Compendium*. John Wiley and Sons, New York, pp. 219–305.
- Robert, S., Bichet, A., Grandjean, O., Kierzkowski, D., Satiat-Jeunemaitre, B., Pelletier, S., Hauser, M.-T., Hofte, H. and Vernhettes, S. (2005) An *Arabidopsis* endo-1,4- β -D-glucanase involved in cellulose synthesis undergoes regulated intracellular cycling. *Plant Cell* 17: 3378–3389.
- Roberts, E.M., Nunna, R.R., Huang, J.Y., Trolinder, N.L. and Haigler, C.H. (1992) Effects of cycling temperatures on fiber metabolism in cultured cotton ovules. *Plant Physiol.* 100: 979–986.
- Rowland, S.P. and Bertoniere, N.R. (1985) Chemical methods for studying supramolecular structure. In: T.P. Nevell and S.H. Zeronian (Eds.), *Cellulose Chemistry and its Applications*, Ellis Horwood, Chichester, pp. 112–137.
- Ruan, Y.-L. (2007) Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fiber. *Func. Plant Biol.* 34: 1–10.
- Ryser, U. (1985) Cell wall biosynthesis in differentiating cotton fiber. *Eur. J. Cell Biol.* 39: 236–256.
- Ryser, U. (1999) Cotton fiber initiation and histodifferentiation. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 1–46.
- Salnikov, V., Grimson, M.J., Seagull, R.W. and Haigler, C.H. (2003) Localization of sucrose synthase and callose in freeze substituted, secondary wall stage, cotton fibers. *Protoplasma* 221: 175–184.
- Saxena, I.M. and Brown, R.M. (2005) Cellulose biosynthesis: Current views and evolving concepts. *Ann. Bot.* 96: 9–21.
- Scheible, W.-R. and Pauly, M. (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr. Opin. Plant Biol.* 7: 1–11.
- Schneider, M., Schwart, J.P.J. and Sanderson, R.D. (1996) Biological variation in the degree of polymerization of cotton lint cellulose produced in South Africa Part II: Comparing Acala 1517 of the Middle Transvaal and the Northern Cape. *Text. Res. J.* 66: 428–435.

- Schrick, K., Fukioka, S., Takatsuto, S., Stierhof, Y.-D., Stransky, H., Yoshida, S. and Jurgens, G. (2004) A link between sterol biosynthesis, the cell wall, and cellulose in *Arabidopsis*. *Plant J.* 38: 227–243.
- Seagull, R.W. (1993) Cytoskeletal involvement in cotton fiber growth and development. *Micron* 24: 643–660.
- Shi, Y.-H., Zhu, S.-W., Mao, X.-Z., Feng, J.-X., Qin, Y.-M., Zhang, L., Cheng, J., Wei, L.-P., Wang, Z.-Y. and Zhu, Y.-X. (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18: 651–664.
- Smyth, G. K. and Speed, T. P. (2003). Normalization of cDNA microarray data. *Methods* 31: 265–273.
- Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3(1): Article 3.
- Smyth, G. K. (2005). Limma: linear models for microarray data. In: R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (Eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Springer, New York, pages 397–420.
- Smyth, G. K., Michaud, J. and Scott, H. (2005). The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067–2075.
- Soltis, P.S., Endress, P.K., Chase, M.W. and Soltis, D.E. (2005) *Phylogeny & Evolution of Angiosperms*. Sinauer Associates, Inc.: Sunderland, p. 190.
- Somerville, C. (2006) Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22: 53–78.
- Stephens, S.G. (1970) The botanical identification of archaeological cotton. *Amer. Antiquity* 35 (3): 368–373.
- Stone, B. (2001) Cellulose: Structure and distribution, in *Encyclopedia of Life Sciences*, Nature Publishing Group, pp. 1–9.
- Szyjanowicz, P.M.J., McKinnon, I., Taylor, N.G., Gardiner, J., Jarvis, M.C. and Turner, S.R. (2004) The *irregular xylem 2* mutant is an allele of korrikan that affects the secondary cell wall of *Arabidopsis thaliana*. *Plant J.* 37: 730–740.
- Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K. and Turner, S. R. (2003) Interactions among three distinct Cesa proteins essential for cellulose synthesis, *Proc. Natl. Acad. Sci. USA* 100: 1450–1455.
- Timpa, J.D. (1991) Application of universal calibration in gel permeation chromatography for molecular weight determination of plant cell wall polymers: Cotton fiber. *J. Agric. Food Chem.* 39: 270–275.
- Timpa, J.D. and Triplett, B.A. (1993) Analysis of cell-wall polymers during cotton fiber development. *Planta* 189: 101–108.
- Timpa, J.D. and Ramey, H.H. (1989) Molecular characterization of three cotton varieties. *Text. Res. J.* 59: 661–664.
- Timpa, J.D. and Ramey, H.H. (1994) Relationship between cotton fiber strength and cellulose molecular weight distribution: HVI calibration standards. *Text. Res. J.* 64: 557–562.
- Truernit, E., Siemerling, K.R., Hodge, S., Vojislava, G. and Haseloff, J. (2006) A map of KNAT gene expression in the *Arabidopsis* root. *Plant Mol. Biol.* 60: 1–20.
- Tu, L.-L., Zhang, X.-L., Liang, S.-G., Liu, D.-Q., Zhu, L.-F., Zeng, F.-C., Liu, D.-Q., Zhu, L.-F., Zeng, F.-C., Nie, Y.-C., Guo, X.-P., Deng, F.-L., Tan, J.-F. and Xu, L. (2007) Gene expression analysis of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep* DOI 10.1007/s00299-007-0337-4.
- Ubeda-Tomas, S., Edvardsson, E., Eland, C., Singh, S.K., Zadik, D., Aspeborg, H., Gorzsas, A., Teeri, T.T., Sundberg, B., Persson, P., Bennett, M. and Marchant, A. (2007) Genomic-assisted identification of genes involved in secondary growth in *Arabidopsis* utilizing transcript profiling of poplar wood-forming tissues. *Physiol. Plant.* 129: 415–428.
- Udall, J.A., Swanson, J.M., Haller, K., Rapp, R.A., Sparks, M.E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A.B., Sickler, B.A., Wilkins, T.A., Guo, J.Y., Chen, X.Y.,

- Scheffler, J., Talierco, E., Turley, R., McFadden, H., Payton, P., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S.R., Pierce, M.L., Essenberg, M., Kim, H., Llewellyn, D.J., Dennis, E.S., Kudrna, D., Wing, R., Paterson, A.H., Soderlund, C. and Wendel, J.F. (2006) A global assembly of cotton ESTs. *Gen. Res.* 16: 441–50.
- Wang, J., Howles, P.A., Cork, A.H., Birch, R.J. and Williamson, R.E. (2006) Chimeric proteins suggest that the catalytic and/or C-terminal domains give CesA1 and CesA3 access to their specific sites in the cellulose synthase of primary walls. *Plant Physiol.* 142: 685–695.
- Wendel, J.F. and Cronn, R.C. (2002) Polyploidy and evolutionary history of cotton. *Adv. Agron.* 78: 139–186.
- Whittaker, D.J. and Triplett, B.A. (1999) Gene-specific changes in alpha-tubulin transcript accumulation in developing cotton fibers. *Plant Physiol.* 121: 181–188.
- Willison, J.H.M. and Brown, R.M. (1977) An examination of the developing cotton fiber: wall and plasmalemma. *Protoplasma* 92: 21–41.
- Wu, Y.-T. and Liu, J.-Y. (2005) Molecular cloning and characterization of a cotton glucuronosyltransferase gene. *J. Plant Physiol.* 162: 573–582.
- Yatsu, L.Y. (1983) Morphological and physical effects of colchicine treatment on cotton (*Gossypium hirsutum* L.) fibers. *Text. Res. J.* 53: 515–519.
- Yuen, C.Y.L., Sedbrook, J.C., Perrin, R.M., Carroll, K.L. and Masson, P.H. (2005) Loss-of-function mutations of ROOT HAIR DEFECTIVE3 suppress root waving, skewing, and epidermal cell file rotation in Arabidopsis. *Plant Physiol.* 138: 701–714.
- Zhang, D., Choi, D.W., Wanamaker, S., Fenton, R.D., Chin, A., Malatrasi, M., Turuspekoy, Y., Walia, H., Akhunov, E.D., Kianian, P., Otto, C., Simons, K., Deal, K.R., Echenique, V., Stamova, B., Ross, K., Butler, G.E., Strader, L., Verhey, S.D., Johnson, R., Altenbach, S., Kothari, K., Tanaka, C., Shah, M.M., Laudencia-Chingcuanco, D., Han, P., Miller, R.E., Crossman, C.C., Chao, S., Lazo, G.R., Klueva, N., Gustafson, J.P., Kianian, S.F., Dubcovsky, J., Walker-Simmons, M.K., Gill, K.S., Dvorak, J., Anderson, O.D., Sorrells, M.E., McGuire, P.E., Qualset, C.O., Nguyen, H.T. and Close, T.J. (2004a) Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L.). *Genetics* 168: 595–608.
- Zhang, D., Hrmova, M., Wan, C.-H., Wu, C., Balzen, J., Cai, W., Wang, J., Densmore, L.D., Fincher, G.B., Zhang, H. and Haigler, C.H. (2004b) Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. *Plant Mol. Biol.* 54:353–372.
- Zhong R., Kayes, S. J., Schroeder, B. P. and Ye, Z.H. (2002) Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. *Plant Cell* 14: 165–179.
- Zhong, R., Richardson, E.A. and Ye, Z.-H. (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. *Planta* 225: 1603–1611.

Responses of the Cotton Genome to Polyploidy

Keith L. Adams, Lex Flagel, and Jonathan F. Wendel

Abstract Cotton has been developed as a particularly useful model system for examining the responses of the genome to polyploidy. Recent studies have provided novel insights and perspectives on genome-wide consequences of polyploidy, as well as the responses of individual genes and effects on gene expression. Comparative BAC sequencing has revealed evidence of genome downsizing in *G. hirsutum*. Expression studies have shown extensive organ-specific gene silencing and expression changes upon allopolyploidy that continue over evolutionary time. Expression patterns can be partitioned between homoeologous genes such that only one copy is expressed in some organs and only the other copy is expressed in other organs. Abiotic stress can have major effects on the expression of homoeologous genes. Larger scale approaches are starting to be employed that discriminate expression of thousands of homoeologs in a single experiment. An improved understanding of the genomic responses to polyploidy will likely have implications and applications for cotton crop improvement.

1 Introduction

Because of its economic importance and the consequent long history of scientific study, the genus *Gossypium* has become one of the best systems for studies of polyploidy. A single polyploidy event occurred during the evolution of the genus, approximately 1.5 mya (Senchina et al. 2003), between a D-genome species closely related to *G. raimondii* and an A-genome species much like modern *G. arboreum* and *G. herbaceum* (Figure 1). This event subsequently led to the evolution of the five species recognized today, including the two domesticated species *G. hirsutum* and *G. barbadense*, and three exclusively wild species, i.e.,

K.L. Adams (✉)

Department of Botany and UBC Botanical Garden and Centre for Plant Research,
University of British Columbia, 6270 University Blvd., Vancouver, BC, V6T 1Z4,
Canada

e-mail: Keitha@interchange.ubc.ca

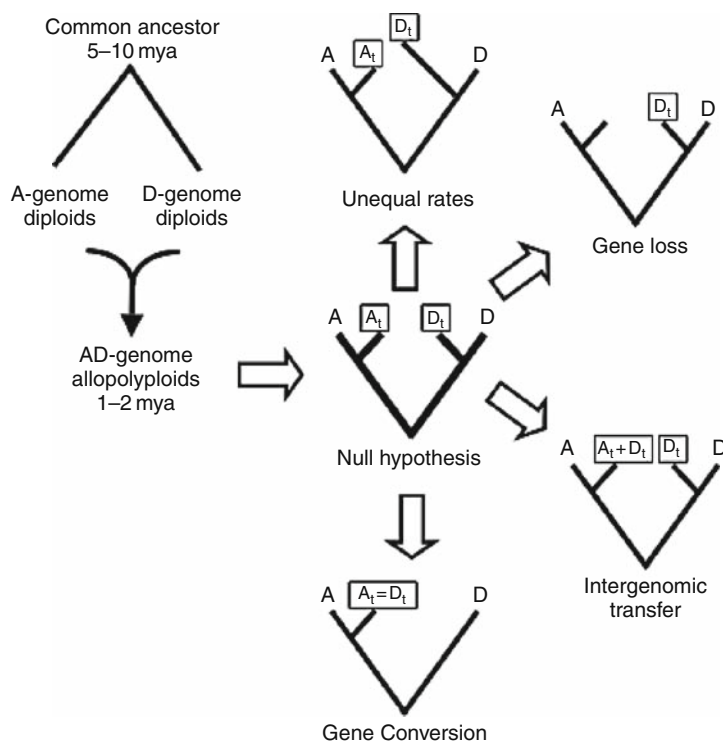


Fig. 1 Phylogenetic history of diploid and allopolyploid *Gossypium* species (left) and various possibilities for subsequent gene and genome evolution (right). At formation, allopolyploids are expected to have duplicated copies (A_t and D_t) of most single-copy and low-copy genes, and duplicated suites of repetitive DNAs. In the absence of mutation or selection, homoeologous copies are expected to evolve at equivalent rates and independently of one another, such that they are phylogenetically sister to their counterparts from the progenitor diploids rather than to each other (center). This expectation provides a convenient null hypothesis for diagnosing molecular evolutionary phenomena that accompany genome doubling, such as unequal rates of sequence evolution, gene loss, intergenomic transfer, and gene conversion

G. darwinii, *G. tomentosum*, and *G. mustelinum*. For more on the evolution and natural history of the cotton genus, see Wendel et al. (2007) in this volume.

In this chapter we focus on the myriad genomic consequences of the genome merger and chromosome doubling that gave rise to the allopolyploid cottons. We consider responses that have occurred on an evolutionary time scale since polyploid formation, as well as those that arise immediately or within a few generations after polyploidy, the latter detected by the use of synthetic cotton neopolyploids. We have grouped the genetic and genomic effects of polyploidy into three categories: (1) genome-wide processes, (2) individual gene fates, and (3) gene expression changes. Each of these categories is discussed in turn.

2 Genome-wide responses to polyploidy

2.1 Perspectives on Polyploid Cotton Genome Evolution from BAC Sequencing

Recent sequencing of bacterial artificial chromosomes (BACs) from *Gossypium* species shed light on how the *Gossypium* genome has evolved in polyploid and diploid species. Two comparative BAC sequencing studies have been published: A large region surrounding the *CesA1* gene (Grover et al. 2004) from both homoeologous genomes of *G. hirsutum*, and a large region surrounding the *AdhA* gene (Grover et al. 2007) from *G. hirsutum* and diploids representing models of the two progenitor genomes, i.e., *G. arboreum* and *G. raimondii*.

Analyses of the *AdhA* BAC region between the diploids and polyploids indicated that more small deletions are present in both polyploid genomes (designated A_t and D_t) than in either diploid genome, illustrative of the general phenomenon of genomic downsizing in polyploid genomes (Bennett and Leitch 2005), while providing a glimpse into one of the underlying mechanisms. Also, the total length of transposable elements was less in the A_t and D_t genomes compared with the A and D genomes. Some *gypsy* elements and intergenic regions show evidence of illegitimate recombination that has led to small deletions. Presently, it is difficult to evaluate genome evolution since polyploidy in the *CesA1* region because BACs from the diploids have not been sequenced. Ongoing sequencing of additional homologous and homoeologous BAC-sized regions of the genome from the diploids and polyploid, will shed additional light on features of genome evolution that distinguish polyploid cotton from its diploid progenitors.

2.2 Movement of Repetitive Elements

Some studies have shown that dispersed repetitive elements have become mobilized as a consequence of polyploidization in cotton, raising the possibility that new insertions have created novel regulatory changes or gene functions. The studies of Zhao et al. (1998) and Hanson et al. (1998) are noteworthy in this respect. Using florescent *in situ* hybridization, they showed that dispersed repetitive sequences that are A-genome-specific at the diploid level have colonized the D-genome at the polyploid level. Similarly, Hanson et al. (1999) showed that a family of *copia*-like retro-transposable elements “horizontally” transferred across genomes following allopolyploid formation. These and other studies highlight the evolutionary possibility of transposable element spread across genomes following polyploid formation, and raise the possibility that this process has played a role in diversification and adaptation.

2.3 Genomic Stasis upon Polyploidy

In contrast to the studies of the natural polyploid *Gossypium* species discussed above, newly created synthetic cotton polyploids show relative stasis at the DNA level. Liu, Brubaker, Cronn, and Wendel (2001) used AFLP analysis to evaluate the extent of fragment additivity in nine sets of newly synthesized allotetraploid and allohexaploid *Gossypium*. Though approximately 22,000 genomic loci were examined, fragment additivity was observed in nearly all cases, even when methylation-sensitive and insensitive isoschizomers were used. These indications of genomic additivity and epigenetic stasis during allopolyploid formation provide a contrast to evidence from several other model plant allopolyploids, most notably wheat (Feldman et al. 1997; Liu et al. 1998a; Liu et al. 1998b; Ozkan, Levy, and Feldman 2001; Shaked et al. 2001) and *Brassica* (Song, Liu, Tang, and Osborn 1995; Lukens et al. 2006), where rapid and extensive genomic changes including changes in DNA cytosine methylation have been reported.

3 Fates of genes duplicated by polyploidy

3.1 Bi-Directional Concerted Evolution of rDNA Genes

Wendel, Schnabel, and Seelenan (1995) demonstrated interaction among the 18S-26S ribosomal genes that exist at multiple loci in the A- and D-genomes. Specifically, instead of evolving independently, as would be expected if allopolyploid sub-genomes did not interact, repeats at the different loci in allopolyploid cotton become homogenized to the same sequence (either “A-like” or “D-like”) by one or more processes of concerted evolution (reviewed by Elder et al. 1995; see Figure 1). In four of the five allopolyploid species, interlocus homogenization has created exclusively D-genome like rDNAs, whereas in *G. mustelinum* nearly all rDNA repeats have been homogenized to an A-like form. This example showed that since polyploid formation about 1.5 mya some 3800 repeats, each approximately 10 kb in length, were “overwritten” with the alternative form originating from the other parental genome, probably through unequal crossing over or gene conversion. Moreover, interlocus concerted evolution was bi-directional, operating in different directions in different allopolyploid lineages. Interlocus concerted evolution of rDNA has also been documented in other plants, including *Nicotiana* (Lim et al., 2004) and *Glycine* (Joly et al. 2004). What mechanisms underlie rDNA interactions? Unequal crossing-over and gene conversion are likely possibilities, and inter-chromosomal exchanges are probably facilitated by the near-telomeric location of the rDNAs in some plants including *Gossypium* (Wendel, 2000).

In an analogous study, Cronn, Zhao, Paterson, and Wendel (1996) showed that the duplicated arrays of tandemly repeated 5S rDNA genes are not homogenized by concerted evolutionary forces in the allopolyploid, in contrast to the 18S-26S arrays. Similarly, low-copy nuclear genes duplicated by allopolyploidy largely evolve independent of one another in the polyploid nucleus (Cronn, Small, and Wendel 1999; Senchina et al. 2003). To date there has been no convincing demonstration of interlocus gene conversion for single-copy or low-copy nuclear genes in cotton.

3.2 Sequence Rate Acceleration in the D-Genome

A direct test of the null hypothesis of rate equivalence for homoeologous genes is provided by measures of nucleotide diversity levels. If evolutionary forces are equivalent for duplicated genes, mutations should accumulate randomly with respect to homoeolog, and hence in a survey of allelic polymorphism in a sample of individuals, the number of alleles detected should be approximately equal for the two gene copies. This was the approach used by Small et al. (1999) in a study of approximately 1 kb of *AdhA* sequence for 22 accessions (44 alleles per genome) of *G. hirsutum* and for five accessions (10 alleles per genome) of *G. barbadense*. In both allopolyploid species, estimates of nucleotide diversity were higher for *AdhA* from the D-genome than from the A-genome, by a factor of two or greater. In a follow-up study wherein a 1.3 kb section of a second alcohol dehydrogenase gene (*AdhC*) with a faster overall evolutionary rate was sequenced, the same conclusion was even more strongly supported (Small et al. 2002). In a survey of 44 alleles from each genome of *G. hirsutum*, 24 different alleles were detected for the D-genome homoeolog versus only 7 for the A-genome homoeolog. To evaluate whether this was a species-specific effect, 12 alleles were sequenced from each genome of a second allopolyploid species, *G. barbadense*. Although diversity levels were lower, the same phenomenon of differential diversity was observed, with 3 and 1 alleles detected for the D- and A-genome homoeologs, respectively.

The preceding observations suggest that there has been an overall acceleration in evolutionary rate in the D_T-genome relative to the A_T-genome of allopolyploid *Gossypium* (Figure 1). Although this rate enhancement is not always observed (Cronn et al. 1999; Senchina et al. 2003), the emerging picture is that evolutionary forces operating on the two genomes may be fundamentally different. At present, the responsible forces and underlying molecular mechanisms are obscure, but a logical suggestion is that they are causally connected to the nearly two-fold difference in genome size between the co-resident genomes in the allopolyploid nucleus (the A_T genome is approximately twice the size of the D_T genome).

4 Expression Changes and Silencing of Genes Duplicated by Polyploidy

4.1 Organ-Specific Expression Biases of Homoeologous Genes in *G. hirsutum*

In addition to evolutionary changes in gene and genome structure, a key component of polyploid genome evolution concerns the consequences of genome doubling on gene expression (see Figure 2). Adams, Cronn, Percifield, and Wendel (2003) examined the expression levels of a set of 40 homoeologous genes in various organs and genotypes of polyploid cotton. They found that almost one-third of the genes examined in *G. hirsutum* revealed appreciable bias toward one homoeolog or the other in at least one organ. Transcript levels for the two members of each gene pair varied considerably by gene and, unexpectedly, by organ. Floral organs showed particularly interesting expression patterns in this regard, with major differences among stamens, petals, and stigmas/ styles. Particularly noteworthy were genes such as the alcohol dehydrogenase gene *AdhA* that showed organ-specific, reciprocal silencing of alternate homoeologs, where there is minimal to no transcription of one member of a duplicated gene pair in some organs and a similar absence of expression of its duplicate in other parts of the plant. Not only can homoeologous gene expression patterns vary in different plant organs but there can be extensive variation

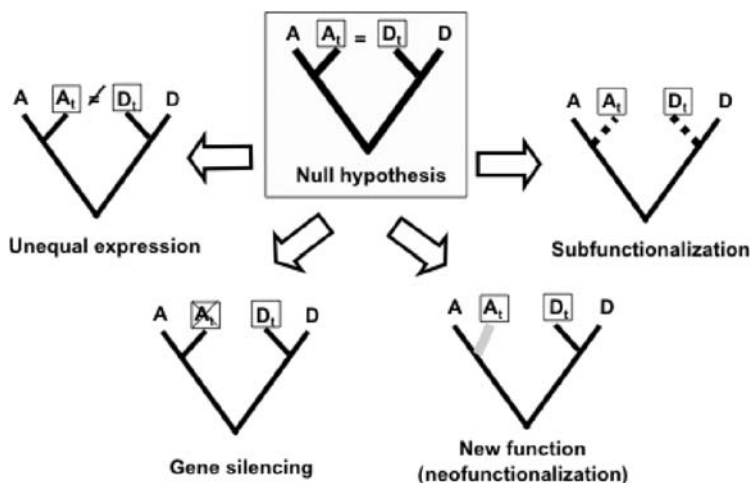


Fig. 2 A model of various possibilities for gene expression and functional evolution after polyploidy. The null hypothesis for a gene that was expressed at equal levels in the diploid parents is that both homoeologs (A_1 and D_1) are equally expressed and retain the same function. However mutation or selection may cause various types of alterations in gene expression and function: unequal expression, gene silencing, neofunctionalization (gain of a new function), or subfunctionalization

during the development of a particular organ. Liu and Adams (2007) showed that the A_T/D_T ratio of *AdhA* homoeologs changed considerably during the development of hypocotyls and cotyledons at 8 stages of development as well as in ovary walls during 11 developmental stages.

The process of subfunctionalization, partitioning of function and/or expression pattern between two duplicate genes, is thought to be one that operates on an evolutionary timescale, requiring fixation of complementary degenerate mutations in regulatory regions or functional domains (Force et al. 1999; Lynch and Force 2000). The data discussed above suggest that for a few genes such as *AdhA*, homoeologous gene pairs have undergone subfunctionalization of expression patterns during the evolution of the cotton polyploids in the last 1.5 million years. These examples from cotton probably represent the most evolutionarily recent examples of subfunctionalization reported in plants.

To study homoeologous gene expression patterns in *G. hirsutum* at a much larger scale than was done with the SSCP-cDNAs assays used by Adams et al. (2003), Udall et al. (2006) developed homoeolog-specific microarrays (see chapter by Udall and Wendel, this volume). The probes on the arrays were developed based on single nucleotide polymorphisms between the A_T and D_T genomes of the allopolyploid. Their arrays contained diagnostic probes for 461 genes. Using leaf RNA, Udall et al. (2006) found that 199 out of 461 genes (43%) showed deviations from equal expression. Expression levels of five genes were confirmed by SSCP-cDNA and the results were similar to the microarray results. Ongoing studies will provide additional insights into the extent of expression biases for hundreds of homoeologous gene pairs.

The findings of expression biases for one homoeolog or the other in allopolyploids that varies by organ type raise questions as to its biological meaning. Biased expression levels might reflect the contributions of each homoeolog to the transcriptome in all cell types within an organ. Alternatively, because organs are composed of multiple cell types, the observation of biased expression could be due to silencing of one homoeolog in some cell types and expression in others within the organ. Ongoing work in the Wendel lab has found that the proportion of the transcript pool contributed by homoeologs, for hundreds of genes, can vary during development of a single cell, specifically, during fiber cell development. Collectively, the above results show that gene expression is massively altered in polyploid cotton relative to its diploid progenitors, and that the specifics of expression alteration vary in different parts of the plant, among genes, and can be developmentally regulated in heretofore unanticipated ways.

4.2 Effects of Abiotic Stress Conditions on Homoeologous Gene Expression

Could environmental stress affect homoeologous gene expression? To begin to answer this question Liu and Adams (2007) examined expression of *AdhA* homoeologs under three types of abiotic stress conditions, including cold,

dark, and water submersion, at different developmental stages. They found that some stress treatments significantly altered the ratio of A_T to D_T expression. Most notably, only one copy of *AdhA* is expressed in hypocotyls during a water submersion treatment and only the other copy is expressed during cold stress. These results imply that subfunctionalization of genes duplicated by polyploidy has occurred in response to abiotic stress conditions, an unprecedented finding. Partitioning of duplicate gene expression in response to environmental stress may lead to duplicate gene retention during subsequent evolution. It will be interesting in the future to examine expression patterns of other genes in response to abiotic stress conditions to determine the prevalence of this phenomenon.

4.3 Immediate Effects of Allopolyploidy on Gene Expression

Do changes in gene expression occur immediately on allopolyploidy or are the effects mostly seen on a long term evolutionary time scale? To answer this question Adams, Percifield, and Wendel (2004) studied a first generation synthetic allopolyploid with genomic composition AAGG, formed by hybridization between *G. arboreum* and *G. bickii* followed by chromosome doubling (Brubaker, Paterson, and Wendel 1999). They surveyed about 2000 loci, using AFLP-cDNA analysis, and found that about 5% of the homoeologous genes were silenced or down-regulated in the polyploid compared with its parents. Genes for a variety of functions were silenced in a homoeolog-specific manner (i.e., either the maternal or paternal copy). Comparisons of expression patterns among eight organs in the allopolyploid, with a focus on floral organs, showed that silencing and preferential expression of one copy are organ-specific, as was seen in *G. hirsutum*. These results indicate that homoeologous gene silencing can occur immediately upon allopolyploid formation in an organ-specific manner.

How repeatable is homoeologous gene silencing among independently created allopolyploids? Adams et al. (2004) found that a monooxygenase gene showed silencing of the maternal A-genome copy in three independently created neopolyploid cottons of different parental origin (Adams et al. 2004). In addition there was similar organ-specific silencing and biased expression of *AdhA* homoeologs in two AADD synthetic allotetraploids as well as in *G. hirsutum* (Adams et al. 2003; Adams 2008). Not all genes, however, show repeatable silencing patterns; two other genes investigated by Adams et al. (2004) showed variable expression and silencing patterns among three synthetic allopolyploids.

Is the homoeologous gene silencing seen in the synthetic cotton allopolyploids caused by hybridization between two species, or by doubling of the chromosomes? Can interspecific hybridization by itself cause gene silencing, in particular silencing that is organ-specific? To gain insights into these questions Adams and Wendel (2005) examined expression of the *AdhA* gene in two

different cotton F₁ hybrids: *G. raimondii* x *G. gossypoides* and *G. trilobum* x *G. gossypoides*. One copy of *AdhA* was silenced in both hybrids. Most notably there was reciprocal silencing in the *G. raimondii* x *G. gossypoides* hybrid such that only the allele from *G. gossypoides* was expressed in some organs and only the *G. raimondii* allele was expressed. This phenomenon parallels the organ-specific silencing of homoeologs in the allopolyploids except that it involves alleles instead of homoeologs.

Ongoing work in the Wendel lab compared expression between a diploid F₁ hybrid that formed by hybridization between an A-genome diploid (*G. arboreum*) and a D-genome diploid (*G. raimondii*) with the natural allotetraploid *G. hirsutum*, using the aforementioned homoeolog-specific microarray platform (Flagel et al. 2008). By comparing the homoeolog expression biases between these species, Flagel et al. can isolate those expression biases that are contributed by genomic merger from the biases caused by subsequent evolution in the natural allopolyploid. In petal tissues, it has been determined that among approximately 1200 genes, the F₁ hybrid shares an expression bias with the allopolyploid about 41% of the time. This intriguing result indicates that genomic merger alone contributes significantly and immediately to homoeolog expression. On the other hand, a still larger fraction (59%) of homoeolog expression biases were found not to be conserved between the F₁ hybrid and allopolyploid. Thus, the effects of genome doubling and 1.5 million years of duplicate gene expression evolution play a larger role in shaping homoeolog expression patterns than does genome merger. From these results it can be concluded that both immediate and long-term evolutionary factors contribute to the creation of homoeologous expression patterns, positing that homoeolog expression biases are induced in *Gossypium* allopolyploids in two distinct phases. Substantial change appears to occur immediately upon allopolyploid formation and within a few generations, followed by slower processes of evolutionary tinkering permitted by duplicate genes.

5 Perspectives

Our understanding of the responses of the cotton genome to polyploidy have made great strides in the past few years. Particularly noteworthy advances have been made with comparative BAC sequencing and expression studies, the latter now being examined using high-throughput methods such as custom microarrays (see chapter by Udall, this volume). Future studies using these and other technologies will provide a broader perspective on the extent and patterns of gene expression changes and gene silencing in allopolyploid cotton. Follow-up functional and transgenic experiments will be required for additional insight into the possible physiological and evolutionary significance of the various patterns of duplicate gene expression found in allopolyploid cotton.

Acknowledgments We thank a reviewer for helpful comments on the manuscript. Research on cotton genome and gene expression evolution in the Wendel lab has been funded by the NSF Plant Genome and the USDA NRI programs, whose support we gratefully acknowledge.

References

- Adams K.L., Cronn R., Percifield R., and Wendel J.F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA* 100, 4649–4654.
- Adams K.L., Percifield R., and Wendel J.F. (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217–2226.
- Adams K.L. and Wendel J.F. (2005). Allele-specific, Bi-directional Silencing of an Alcohol Dehydrogenase Gene in Different Organs of Interspecific Diploid Cotton Hybrids. *Genetics* 171, 2139–2142.
- Adams K.L. (2008) Insights into the evolution of duplicated gene expression in polyploids from *Gossypium*. *Botany* 86, 827–834.
- Bennett, M.D. and Leitch, I.J. (2005) Genome size evolution in plants. In: T. R. Gregory (Ed.) *The evolution of the genome*. Elsevier, San Diego, pp. 89–162.
- Brubaker, C. L., Paterson, A. H., and Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42, 184–203.
- Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705.
- Cronn, R., Small, R. L., and Wendel, J. F. (1999). Duplicated genes evolve independently following polyploid formation in cotton. *Proc. Natl. Acad. Sci. USA* 96, 14406–14411.
- Elder, J. F., and Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart. Rev. Biol.* 70, 297–320.
- Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A., and Vega, J. M. (1997). Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* 147, 1381–1387.
- Flagel L., Udall J., Nettleton D., Wendel J.F. (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 6(16).
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Grover C.E., Kim H., Wing R.A., Paterson A.H., and Wendel J.F. (2007) Microcolinearity and genome evolution in the *AdhA* region of diploid and polyploid cotton (*Gossypium*). *Plant J.* 50, 995–1006.
- Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF. (2004) Incongruent patterns of local and global genome size evolution in cotton. *Genome Res.* 14, 1474–1482.
- Hanson, R. E., Islam-Faridi, M. N., Crane, C. F., Zwick, M. S., Czeschin, D. G., Wendel, J. F., McKnight, T. D., Price, H. J., and Stelly, D. M. (1999). Ty1-*cop* retrotransposon behavior in a polyploid cotton. *Chromosome Res.* 8, 73–76.
- Hanson, R. E., Zhao, X.-P., Islam-Faridi, M. N., Paterson, A. H., Zwick, M. S., Crane, C. F., McKnight, T. D., Stelly, D. M., and Price, H. J. (1998). Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). *Amer. J. Bot.* 85, 1364–1368.
- Joly S, Rauscher JT, Sherman-Broyles SL, Brown AH, Doyle JJ. (2004) Evolutionary dynamics and preferential expression of homoeologous 18S-5.8S-26S nuclear ribosomal genes in natural and artificial glycine allopolyploids. *Mol Biol Evol.* 21, 1409–1421.

- Lim K.Y., Skalicka K., Koukalova B., Volkov R.A., Matyasek R., Hemleben V., Leitch A.R., Kovarik A. (2004). Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. *Genetics* 166, 1935–1946.
- Liu, B., Vega, J. M., and Feldman, M. (1998a). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* 41, 535–542.
- Liu, B., Vega, J. M., Segal, G., Abbo, S., Rodova, M., and Feldman, M. (1998b). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy non-coding DNA sequences. *Genome* 41, 272–277.
- Liu, B., Brubaker, C. L., G M. Cronn, R. C., and Wendel, J. F. (2001). Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 43, 874–880.
- Liu, Z. and Adams, K. L. (2007) Expression partitioning of genes duplicated by polyploidy in response to abiotic stress and during organ development. *Curr. Biol.*, in press.
- Lukens L.N, Pires J.C, Leon E., Vogelzang R., Oslach L., and Osborn T., (2006). Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiol* 140, 336–348.
- Lynch M., and Force A.G., (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.
- Liu, B. and J. F. Wendel. (2002) Non-Mendelian phenomena in allopolyploid genome evolution. *Current Genomics* 3, 489–506.
- Ozkan, H., Levy, A. A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13, 1735–1747.
- Senchina, D. S., Alvarez, I., Cronn, R. C., Liu, B., Rong, J., Noyes, R. D., Paterson, A. H., Wing, R. A., Wilkins, T. A., and Wendel, J. F. (2003). Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Biol. Evol.* 20, 633–643.
- Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13: 1749–1759.
- Small, R. L., Ryburn, J. A., and Wendel, J. F. (1999). Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* 16, 491–501.
- Small, R. L., and Wendel, J. F. 2002. Differential evolutionary dynamics of duplicated paralogous *Adh* loci in allotetraploid cotton (*Gossypium*). *Mol. Biol. Evol.* 19, 597–607.
- Song, K., Lu, P., Tang, K., and Osborn, T. C. (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. USA* 92, 7719–7723.
- Udall, J. A., Swanson, J. M., Nettleton, D., Percifield, R. J., and J. F. Wendel, (2006) *Genetics* 173, 1823–1827.
- Wendel, J. F., Schnabel, A., and Seelanan, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* 92: 280–284.
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249.
- Zhao, X.-P., Si, Y., Hanson, R. E., Crane, C. F., Price, H. J., Stelly, D. M., Wendel, J. F., and Paterson, A. H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res.* 8, 479–492.

Comparative Genomics of Cotton and *Arabidopsis*

Junkang Rong and Andrew H. Paterson

Abstract Upland cotton (*G. hirsutum*), a tetraploid species, evolved from a reunion of two diploid species (most probably *G. herbaceum* and *G. raimondii* with A and D genomes respectively) about 1-2 million year ago. Cytogenetic and molecular evidence indicated that A and D diploid cottons diverged from a common ancestor about 7-10 million years ago. High-density genetic linkage maps reveal the homoeologous relationship between A and D chromosomes, and permit the reconstruction of the gene order of their hypothetical common ancestor. Comparative studies with the inferred gene order of the hypothetical cotton ancestor and *Arabidopsis* pre-duplication ancestor indicate that cotton and *Arabidopsis* share a useful level of correspondence of gene order. This syntenic relationship is very helpful in identification and isolation of important cotton genes using *Arabidopsis* sequence information. As an example, the correspondence between some mapped fiber QTLs and some cloned fiber related cDNA/genes is shown. Cotton fibers and *Arabidopsis* trichomes both develop from epidermal cells and are single celled 'organs'. With the aid of the considerable knowledge of *Arabidopsis* trichome development, some cotton homologs of *Arabidopsis* trichome genes have been cloned and shown to be related to fiber development. An eventual sequence for the cotton genome(s) will further advance the ability of cotton researchers to utilize translational genomics approaches to benefit from *Arabidopsis* functional information, also clarifying the consequences of additional ancient duplication in cotton, and perhaps hinting at the identities of genes that contribute to the unique features of cotton.

J. Rong (✉)

The University of Georgia, Plant Genome Mapping Laboratory
e-mail: jkrong@yahoo.com

1 Introduction

Detailed maps of DNA markers reveal the genetic and evolutionary relationships of different chromosomes or genomes through studies of genomic co-linearity - analyzing the alignments (order and distance) of corresponding markers on different chromosomes or chromosome segments. Such research has aroused considerable scientific interest and been carried out in many plant taxa (Paterson et al., 2000; Eckardt, 2001; Paterson et al., 2004b). Since the publication of the first sequenced angiosperm genome, that of *Arabidopsis thaliana*, about 7 years ago (The Arabidopsis Genome Initiative, 2000), the whole rice genome (Goff et al., 2002; Yu et al., 2002) and a draft sequence of the poplar genome (Tuskan et al., 2006) have also been completed, with more in progress including cotton. Comparative analysis at the DNA level is providing new insights into genome organization and evolution of botanical diversity.

An important benefit from these types of studies is the ability to predict the functions of genes in agriculturally-important crops based on their positions (Oh et al., 2002). Comparative genomics can also provide information such as estimates of divergence time and genetic similarity among different species. Because many crops are polyploid or paleopolyploid with relatively large genomes, using a small fully-sequenced genome such as *Arabidopsis* as a reference can facilitate prediction of the functions of genes in these polyploid species based on positional information.

Cotton, the world's leading natural textile fiber, is a promising beneficiary from extensive gene functional information that is accumulating for *Arabidopsis*, the first plant genome to be fully sequenced. Cotton is the nearest well-mapped relative to *Arabidopsis* outside the *Brassicales* (Bowers et al., 2003). We show herein how patterns of gene co-linearity between cotton and *Arabidopsis* provide clues to detect gene position for important cotton traits.

The genetic basis of cotton fiber quality is naturally of much interest. While little is known about the specific genes responsible for variation in cotton fiber quality, a growing set of clues exist from which to identify such genes. Cotton fibers and *Arabidopsis* trichomes have the same botanical origin and may share a similar genetic basis. So, the well known genetics of the *Arabidopsis* trichome (Larkin et al., 2003; Schiefelbein, 2003) may be a valuable guide to dissecting cotton fiber genetics.

A complication in comparative studies is that duplication of the entire genome of *Arabidopsis* has occurred since its divergence from cotton (Vision et al., 2000; Bowers et al., 2003). Evidence from a recently published high density map of the cotton genome (Rong et al., 2004), together with classical data (Muravenko et al., 1998; Wendel and Cronn, 2003) indicated that such an ancient duplication happened in the cotton lineage as well. The comparison of inferred ancestral gene orders between cotton and *Arabidopsis* has shown more co-linearity than the modern genomes alone. This chapter will summarize findings regarding cotton genomic structure, co-linearity and synteny between cotton and *Arabidopsis*, and its application in identification of important cotton genes and cotton breeding.

2 Foundation of Comparative Genomics

It has been estimated that up to 80% of the extant flowering plants are polyploidy (Stebbins, 1966), and diploid-like species, such as rice (Kishimoto et al., 1994; Paterson et al., 2003; Paterson et al., 2004a) and *Arabidopsis* (Kowalski et al., 1994; Paterson et al., 1996; Paterson et al., 2000; Vision et al., 2000; Blanc et al., 2003; Bowers et al., 2003) are now known to be ancient polyploids. Whole genome duplication or large segmental duplication followed by gene loss, functional divergence, gene movement, and chromosome structural changes such as translocation and inversion, collectively forming a process known as diploidization, contribute to eventual restoration of polyploids to diploid-like genome structure after a long period of evolution. Therefore, diploid-like polyploids are referred to as paleopolyploid.

Genetic maps and DNA sequences permit us to identify paleopolyploids based on colinear arrangements of homologous genes, presumed to be preserved since their divergence from a common ancestor. Co-linearity and/or synteny between different genomes defines evolutionarily related regions that have been preserved since speciation, while co-linearity and/or synteny within a genome defines segmental or whole genome duplication. Detection of corresponding regions is important for understanding the evolutionary history of plant genomes and for making deductions about candidate genes, and is a key aspect of comparative genomics.

However, our ability to detect syntenic regions decreases across evolutionary time because of diploidization processes. Complete sequences or high-density genetic maps improve our ability to discern evolutionary relationships between diverged genomic regions (Odland et al., 2006). In addition, the identification of evolutionary relationships can be further improved using the inferred gene order of pre-duplication ancestral genomes as a reference (Bowers et al., 2003; Rong et al., 2005a), as discussed in detail below.

Discerning a ‘signal’ of co-linearity amongst the ‘noise’ of divergence and/or diploidization can be challenging. A robust method of defining synteny in genomes that have disruptions in co-linearity is by the quality of a diagonal that forms in genomic comparisons using an Oxford grid (Cannon et al., 2003; Odland et al., 2006). An Oxford grid is a comparative-mapping tool that displays homologous sequences between two chromosomes on a two-dimensional grid. Several software packages such as AdHoRe (Vandepoele et al., 2002), FISH (Calabrese et al., 2003), DiagHunter (Cannon et al., 2003) and ColinearScan (Wang et al., 2006) has been developed based on the diagonal quality and are being used in comparative genomic analysis.

3 Paleopolyploidy of Diploid Cotton

An important part of understanding the comparative genomics of cotton and *Arabidopsis* is to shed light on genomic and evolutionary events since their divergence from a common ancestor. Diversification, neopolyploidy, and

paleopolyploidy within *Gossypium* have each had important consequences for cotton-*Arabidopsis* comparisons. Paleopolyploidy in the *Arabidopsis* lineage also complicates this comparison, as has rapid restructuring of *Arabidopsis* in the past few million years. In the near future the sequences of *Arabidopsis lyrata* and *Capsella rubella* will permit us to unravel this rapid restructuring (Koch and Kiefer, 2005, Yogeeswaran et al. 2007).

3.1 Diversification of the *Gossypium* Genus

Gossypium, a genus in the tribe *Gossypieae* of the Malvaceae family, contains 45 diploid and 5 allotetraploid species (Fryxell 1992). Diploid *Gossypium* species can be classified into eight different genome types designated A through G, and K, based on their morphological features and geographical origin together with chromosome size and meiotic pairing behavior in interspecific hybrids (Endrizzi et al., 1985; Wendel and Cronn, 2003). All five allotetraploids originated in the New World from an inter-specific hybridization event involving two diploid species closely related to *G. herbaceum* L. (A₁) and *G. raimondii* L. (D₅) (Wendel, 1989) or *G. gossypoides* (Ulbrich) Standley (D₆) (Wendel et al., 1995). The evolutionary history and genetic relationships of cotton genomes at the diploid and tetraploid levels were largely clear many years ago (Endrizzi et al., 1985; Wendel and Cronn, 2003). Aspects of these relationships are reviewed in detail in the Wendel and Konan chapters of this volume.

3.2 Cytogenetic and Molecular Evidence for Neopolyploidy in *Gossypium*

That the tetraploid cottons are the products of a natural cross between two diploid species was suspected as early as 1923 (Nikolajeva, 1923) and 1924 (Denham, 1924), respectively, with the discovery that some *Gossypium* species have twice the number of chromosomes of other species. Longley (1933) deduced that the species with doubled chromosome numbers are from whole genome duplication. Skovsted (1933, 1934) observed that New World cotton species normally have small chromosomes and Old World species have larger ones; and that A genome diploid chromosomes pair with the 13 larger chromosomes (2.25-2.36µm) of tetraploid cottons in interspecific hybrids, leaving 13 smaller chromosomes (1.25-1.45µm) as univalents. Further cytological study reviewed in detail in the Konan chapter of this volume revealed that the New World tetraploid cottons are true allopolyploids containing A and D sub-genomes (Webber, 1935; Skovsted, 1937; Beasley, 1940). Beasley (1940) and Harland (Harland, 1940), showed that artificial allotetraploids synthesized by making crosses between A genome and D genome species have regular bivalent

pairing like normal diploids and can produce fertile hybrids with natural tetraploid cotton. Based on a range of considerations including leaf developmental patterns, seed hairs and vigor of the hybrids together with their chromosome pairing, the two sub-genomes in tetraploid (At and Dt) were eventually accepted by most people to be from *G. herbaceum* and *G. raimondii* (Endrizzi et al., 1985; Wendel and Cronn, 2003), respectively, as detailed in the Wendel and Konan chapters of this volume.

With the emergence of molecular tools during the 1980's, research at the DNA level showed that tetraploid cottons originated within the past 1-2 million years (Wendel, 1989), from a hybridization between A and D genome ancestors that diverged from a common ancestor about 7-10 million years ago (Wendel and Cronn, 2003). Construction of several high density genetic maps with DNA markers has helped clarify the relationships among each A and D chromosome at both the diploid and tetraploid levels. The collinear arrangements of common markers on homologous chromosomes of the A genome species (*G. herbaceum* and *G. arboreum*), D genome species (*G. raimondii* and *G. trilobium*) and tetraploid cotton genomes permit the inference of the gene order of the common ancestor genome (Fig. 1). A consensus map of gene order in the hypothetical ancestor genome has been constructed based on common markers mapped on 'At' (tetraploid chromosomes derived from the A-genome ancestor), 'Dt' and diploid D genomes (Rong et al., 2004) including 3016 loci and a total length of 2324.7 cM.

3.3 Cytogenetic, EST, and Structural Genomic Evidence for Paleopolyploidization about 15-30 Million Years Ago in a *Gossypium* Ancestor

In classical cytogenetic studies, the chromosome number of 26 in diploid cottons was deduced to be a result of paleopolyploidization in the modern 'diploid' lineages (see Konan et al, this volume). Based on the observation that the gene number for allozyme-encoding loci exceeded the expected number in the diploid, it was proposed that the ancestral diploid genomes of modern tetraploid cotton contained segmental duplication or had resulted from the reunion of two more ancient genomes with 6 and 7 chromosomes (Gottlieb, 1982; Wendel and Percival, 1990). Low copy DNA probes from cotton genomic DNA and cDNA, could detect 2.96 and 2.71 restriction fragments in diploid cotton A and D genome, and 4.12 in tetraploids on average (Reinisch et al., 1994), also suggested that many genes in the A and D genomes may have two or more copies, even considering a few multiple bands caused by restriction sites within the probe sequence. Giemsa banding patterns of the A-diploid and tetraploid cotton genomes suggested that the ancestral cotton genome consisted of 7 homologous pairs of chromosomes (Muravenko et al., 1998).

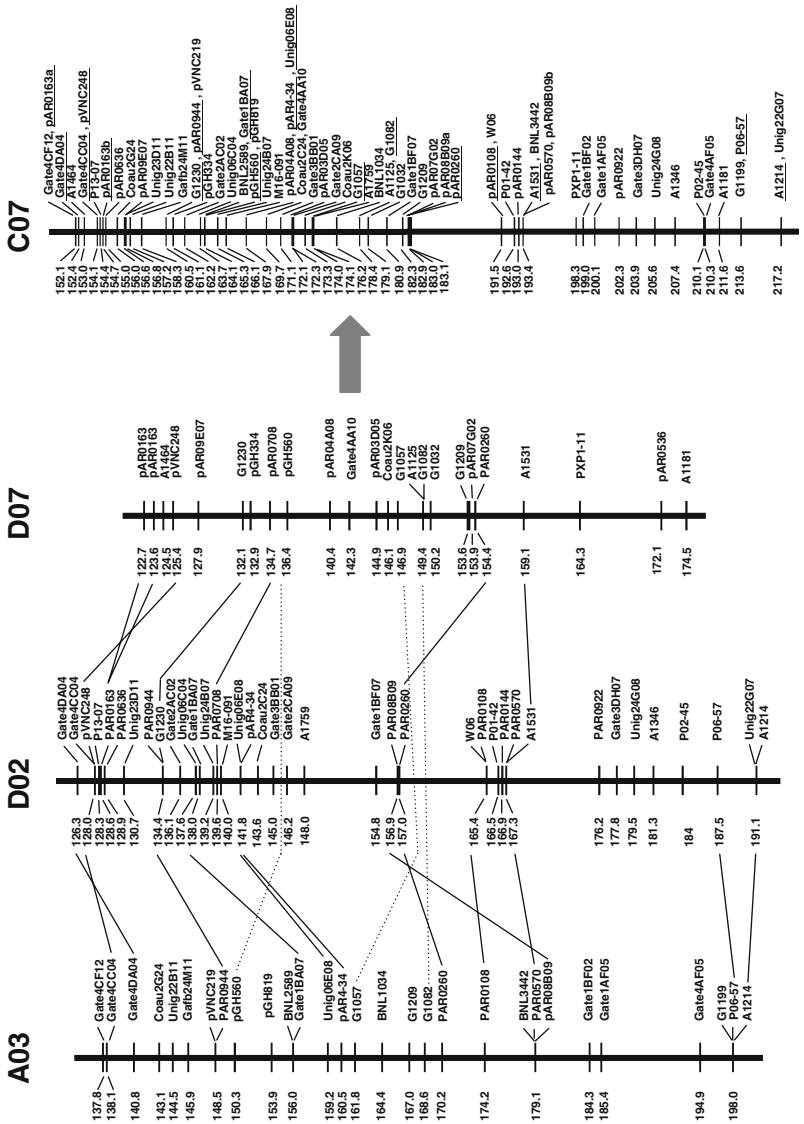


Fig. 1 Genetic maps of partial (bottom 1/3) homologous group 3 (chromosomes LG A03, LG D02 and D2, on the left side of the arrow) and reconstruction of genetic map of their hypothesized ancient chromosome C7 (on the right side of the arrow). Common markers, which were used as anchor loci to build the ancient chromosome (consensus map), were linked with lines in the individual maps and underlined in the consensus map

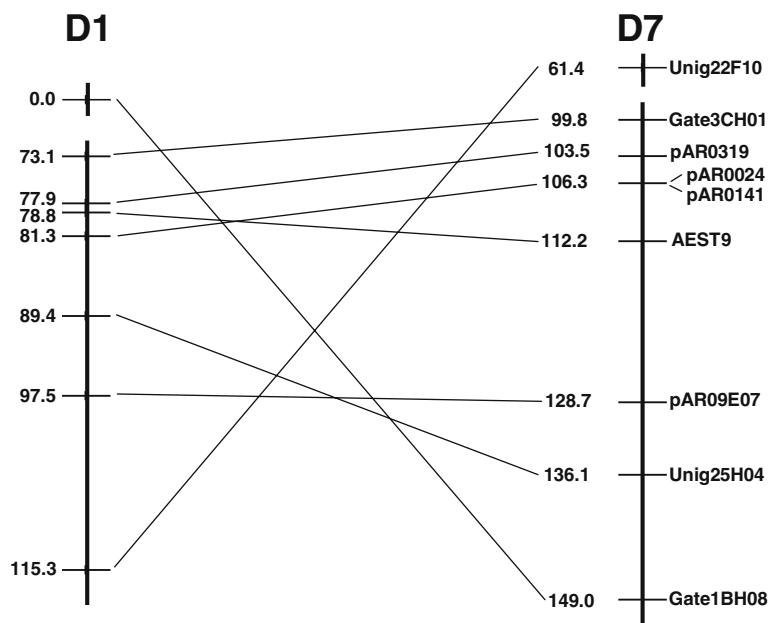


Fig. 2 Two ancient duplicated fragments on chromosomes D1 and 7 of diploid D genome species. Nine duplicated genes were detected on these two chromosomes. Among them, 8 were clustered in a region of 42.2 cM on D1 (total 144.9 cM) except Gate1BH08 far from this region, and 49.2 cM on D7 (total 179.4 cM) except for Unig22F10 with 38.4 cM far from the duplicated region. Revised from Rong et al., 2004

As genetic maps for cotton became more detailed (Brubaker et al., 1999; Rong et al., 2004), it became clear that multiple restriction fragment loci detected by single probes were not randomly distributed on the chromosomes. Duplicate loci mapped in a region of one chromosome were often also clustered on another chromosome within the same diploid or tetraploid subgenome (Fig. 2) (Rong et al., 2004). As a result, each chromosome can be divided into several putatively syntenic blocks (Rong et al., 2004). When the gene order of the hypothetical ancestor genome inferred as shown (Fig 1) was analyzed with two computer programs (FISH and CrimStateII), about 24% and 52%, respectively, of the genome showed intra-genomic correspondence (Rong et al., 2005a). These inferences begin to delineate the duplicated blocks resulting from genome duplication in ‘diploid’ cottons.

A large collection of public expressed sequence tags (EST) for cotton (Udall et al., 2006, see Udall Chapter, this volume) provide additional evidence of an intra-genomic genome duplication. It was found that there were two main peaks in the distribution of synonymous substitutions per synonymous site (Ks) between corresponding genes, at around 0.04 and 0.45 in *G. hirsutum* EST unigenes (Blanc and Wolfe, 2004). Complementary analyses of duplicated unigenes from the diploid *G. arboreum*, yielded a similarly shaped distribution

of Ks values with a clear secondary peak at $Ks = 0.40 - 0.45$ (Blanc and Wolfe, 2004; Rong et al., 2005a), suggestive of a large scale duplication about 15-30 Mya, depending on the neutral substitution rate used (Gaut et al., 1996; Koch et al., 2000).

4 Duplication in the *Arabidopsis* Genome

Arabidopsis thaliana, a species in the family *Brassicaceae*, is a small flowering plant related to cabbage and mustard. Because of its small genome size (about 157 million base pairs), small stature and short life cycle, *A. thaliana* has become a very useful model for understanding the genetic, cellular, and molecular biology of flowering plants.

During meiosis, the five chromosomes of *Arabidopsis* form the five bivalents of a typical diploid. However, numerous lines of evidence show that *A. thaliana* has undergone ancient duplication of chromosomes or chromosome segments. The earliest evidence came from genetic mapping with DNA markers where two or more genetic loci were detected by same DNA probes (McGrath et al., 1993; Kowalski et al., 1994). These multiple loci tended to be non-randomly distributed, with those clustered on a region of one chromosome often having counterparts close together on another chromosome, hinting that these segments were derived from the duplication of an ancestral segment.

With the completion of sequencing of the *Arabidopsis* genome (The Arabidopsis Genome Initiative, 2000), a total of 23177 genes, representing 89% of total *Arabidopsis* ones, could be aligned into 26 large segment pairs named "alpha duplicates" (Bowers et al., 2003). The sizes of alpha duplicates varied from 260 ($\alpha 09$) to 2413 ($\alpha 05$) genes. This finding clarified the genome structure and evolutionary history of *Arabidopsis*, and provided the ancestral gene order needed to improve our ability to reveal synteny with other plants such as cotton.

5 Comparative Organization of Cotton and *Arabidopsis* Genomes

5.1 Evolutionary Context of Two Species

The *Malvales* (including cotton) and *Brassicales* (including *Arabidopsis*) are thought to have diverged from a common ancestor about 80 MYA (million years ago, Bowers et al., 2003). During their independent evolution, genetic changes took place including functional divergence, sequence elimination, and gene and segmental rearrangement. Among cotton EST sequences from various sources including *G. arboreum*, *G. raimondii* and *G. hirsutum* cDNA libraries, about 55-70% have homologs in the *Arabidopsis* genome meeting an E-value less than 10^{-5} , indicating that about 30% of cotton genes are either absent from, or no longer recognizable in, *Arabidopsis*.

5.2 Synteny between Cotton and *Arabidopsis*

The finding that about 55-70% of cotton ESTs have homologs in *Arabidopsis*, makes it possible to test if there is any similarity in gene arrangement between cotton and *Arabidopsis* by analyzing the distance and order of homologs in the respective genomes. In a comprehensive analysis using the consensus maps discussed in section 3.2, 95% of the mapped cotton probes (2162 probes and 2800 loci) could be sequenced and compared to *Arabidopsis* genes (Rong et al., 2005a). The distribution of best-matching *Arabidopsis* sequences on the cotton consensus map was not random. About 12.8 % of consecutive cotton loci have homologs on the same *Arabidopsis* alpha-duplication segments, significantly ($P = 3.02 \times 10^{-63}$) more than the 4.4% that could be explained by chance. To explore the extent and distribution of conserved synteny between cotton and *Arabidopsis*, two software packages, Fast Identification of Segmental Homologs (Calabrese et al., 2003) and CrimeStatII (Levine, 2002), were each used to analyze the 5132 pairs of 'potential orthologs' that met a threshold of $E < 10^{-10}$. CS2 and FISH, respectively, detected 187 clusters including 800 pairs of matching loci, and 310 blocks including 715 pairs of matching loci.

To explore the consequences of duplication/diploidization in *Arabidopsis* for genomic comparisons, the inferred ancestral cotton map was compared to both the modern *Arabidopsis* genome, and to inferred pre-alpha-duplication *Arabidopsis* gene orders (Bowers et al., 2003). When CS2 was used for comparison of cotton to the modern *Arabidopsis* genome, 270 (33.8%) of the 800 pairs of matching loci no longer occurred in clusters, reflecting gene loss in modern *Arabidopsis* chromosomes following the alpha event. Using the inferred ancestral *Arabidopsis* gene order increases the detection of conserved synteny. Based on evaluation of matching loci in their distribution along the cotton chromosomes, CS2 detected correspondence over nearly twice as much of the genome as FISH, both in total and along each chromosome. Most regions detected by CS2 were larger than, and inclusive of, the regions detected by FISH, excepting regions determined with only two pairs of matching loci (CS2 requires at least 3 matching pairs for significance). Overall, a total of 1372.1 cM, or 59.0% of the cotton consensus map, showed non-random correspondence, putatively synteny, with at least one *Arabidopsis* α duplicated segment based on CS2. Different cotton chromosomes varied in the portion over which correspondence could be inferred, from 81.1% for C05 to 40.4% for C10. In most cases (81.1% with CS2, 96.8% with FISH), a single genomic region in the hypothetical ancestral cotton map corresponded with only 1-2 *Arabidopsis* duplicated segments. These percentages varied among chromosomes – for example, more than 80% of correspondence on C01 and C10 involved only one *Arabidopsis* α region while on C03 nearly 70% of correspondence involved two or more *Arabidopsis* regions. Evaluation of putative orthologs (as described above) in their distribution along the *Arabidopsis* genome (measured in the number of genes along the segments) also showed much correspondence with cotton. Segments of

Arabidopsis pre- α gene orders covering 12402 genes (53.5%) showed correspondence with the cotton consensus map using CS2. Again, FISH detected lower correspondence (27.8% of the *Arabidopsis* transcriptome) across the whole genome, and on all but three individual pre-alpha segments ($\alpha 01$, $\alpha 17$, $\alpha S07$).

6 Application of Comparative Information for Cotton – *Arabidopsis* Translational Genomics

6.1 Comparative Genomics of *Arabidopsis* Trichome and Cotton Trichome/Fiber as an Example

The *Arabidopsis* trichome is a single celled organ originating from the epidermis of the leaves, stems, and sepals. All phases of its development can be easily observed, and its development can be genetically dissected with little impact on other aspects of plant growth and development. Understanding trichome development may have several practical implications. First, it is morphologically and developmentally similar to the primary economic product of cotton, seedborne epidermal lint fibers. Second, there is a correlation between the presence of trichomes on leaves and stems of cotton and resistance to herbivory by certain insect pests (Johnson, 1975; Mauricio and Rausher, 1997). Because of these unique advantages in its growth, structure and distribution, trichomes provide an excellent system to study the genetic base of cell fate, such as selection of trichome initiation in a field of epidermal cells, cell fate-determination, changes in the cell-cycle mode and cell-shape control.

Trichome development can be classified into several stages including patterning (selection of cells for trichome initiation) and initiation, endoreduplication, branching, expansion growth and maturation (Hulskamp, 2004). A large number of mutations that affect trichome development have been created (Marks, 1997; Hulskamp et al., 1999). Through studies of these mutants, many genes functioning in different developmental stages have been cloned and the genetic mechanisms of trichome development have been clarified. For example, trichome patterning and initiation are under the control of a small group of genes including positive regulators GLABRA1 (GL1), TRANSPARENT TESTA GLABRA1 (TTG1), GLABRA3 (GL3) and ENHANCER OF GL3 (EGL3), and trichome-suppressing genes TRIPTYCHON (TRY), CAPRICE (CPC) and ENHANCER CAPRICE TRIPTYCHON1 (ETC1). Details of the function of each trichome gene and models regulating trichome development have been reviewed (Marks, 1997; Larkin et al., 2003; Hulskamp, 2004; Ramsay and Glover, 2005).

Cotton fibers can be divided into two distinct types, ‘fuzz’ and ‘lint’ (Fryxell, 1963) based on their length and origination. Fuzz fibers are short (<5 mm) and tightly attach to the seed coat, similar to the seed hairs of wild cottons. Economically-important lint fibers are much longer than fuzz and easily

separated from the seed coat during ginning. Lint fibers are further classified based on average length (“staple”) at maturity as: short staple (< 21 mm), *G. arboreum* and *G. herbaceum*; long staple (28–32 mm), *G. hirsutum* (‘upland’); and extra-long staple (>35 mm), *G. barbadense* (‘Pima’, ‘Egyptian’, or ‘Sea Island’). Similar to *Arabidopsis* trichomes, both kinds of fibers are single cells developed from the epidermis of the ovule.

Due to its large genome size [2246 mega bases (Arumuganathan and Earle, 1991)] and difficulties in transformation of cotton, it has been difficult to identify the specific genes which are critical for fiber development and quality using positional approaches. Because of the similarities between cotton fibers and *Arabidopsis* trichome, i.e. both are single cell developed from the epidermis, and cotton is the nearest relative to *Arabidopsis* outside the Brassicales, knowledge of *Arabidopsis* trichome development is a natural guide in efforts to clone cotton fiber genes. So far, many cotton genes have been cloned by the aid of *Arabidopsis* trichome genes and most appear to have similar functions in the two species. One successful example of this approach is the cloning of cotton MYB genes. During the development of *Arabidopsis* trichomes, several MYB genes with R2/R3 repeats (the transcription factors) such as GL1, GL3, and WER, act as positive regulators (Walker et al., 1999), whereas others (TRY and CPC) act as negative ones (Schellmann et al., 2002; Wada et al., 2002). The typical MYB protein contains three functional domains i.e. a basic DNA-binding domain (DBD), an acidic transacting domain and a negative regulatory domain. Because the DBD is a highly conserved domain consisting of two or three imperfect repeats (R1, R2 and R3), it was used to design PCR primers for amplification of cotton MYB genes. So far, about 65 cotton MYB genes have been identified in this way and some were revealed to be related to fiber development (Loguercio et al., 1999; Suo et al., 2003; Hsu et al., 2005). Among them, GhMYB109 is most important because it is specifically expressed in fiber initials and elongating fibers, and is similar to AtWER and AtGL1 in sequence (Suo et al., 2003). Other examples of isolation of cotton fiber related genes with guidance from *Arabidopsis* trichome genes are GhTTG1 and GhTTG3 (Humphries et al., 2005). In *Arabidopsis*, TTG1, containing four WD-40 repeats, controls trichome initiation through formation of an initiation complex with GL1 and GL3. Using a combination of PCR-based and library screening techniques, Humphries et al (2005) isolated and characterized four cotton homologs of AtTTG1. The four cotton sequences fall into two distinct pairs, with one pair (GhTTG1 and GhTTG3) having 80% amino acid identity to AtTTG1, and the other pair (GhTTG2 and GhTTG4) around 62%. All four genes are expressed in many tissues throughout the plant, including ovules and growing fibers.

Some cotton fiber related genes directly affect trichome growth in transgenic *Arabidopsis*. For example, real-time PCR and in situ analysis showed that GaMYB2, a MYB-like gene amplified from *G. arboreum*, is predominantly expressed early in developing cotton fibers (Wang et al., 2004). When its coding region and cDNAs were inserted between the 5'-upstream and 3'-downstream

fragments of *GL1* and transferred into *Arabidopsis*, GL1::GaMYB2 rescued trichome formation of a *gll* mutant (Wang et al., 2004). Interestingly, combination of the GaMYB2 coding region with 35S promoter also induced (albeit limited) seed-trichome production (Wang et al., 2004). This demonstrated that both *GL1* and GaMYB2 play a role in patterning trichomes: *GL1* acts as an enhancer in trichome and a repressor in nontrichome cells, generating a trichome-specific pattern of MYB gene expression. MYB109, amplified from *G. arboreum*, displayed similar function as GaMYB2 in *Arabidopsis* trichome development. In functional testing of the four cotton TTG genes in the *Arabidopsis ttg1* mutant (Humphries et al., 2005), which has a number of defects including an absence of trichomes and seed coat pigments, both GhTTG1 and GhTTG3 were able to complement the *ttg1* mutation and restore normal trichome formation after their stable transformation into *Arabidopsis*. It is tempting to speculate that such WD-repeat proteins could have a pivotal role in cotton fiber initiation. In addition, two cotton lipid transfer protein genes, LTP3 and LTP6, are revealed to be specifically expressed in fiber cells (Hsu et al., 1999; Liu et al., 2000). When constructs of their promoters with GUS genes are transformed into tobacco plants, GUS genes only express in trichomes, suggesting that LTP3 and LTP6 promoters confer trichome-specific expression. These results demonstrate molecular parallels between differentiation of trichomes in cotton and other plants such as *Arabidopsis* and tobacco, toward the functional characterization of regulators of cotton fiber cell growth.

6.2 Positional Correspondence of Fiber/Trichome Genes to Fiber QTLs

Research into the molecular and genetic basis for cotton fiber development can be broadly classified into two approaches, i.e. mapping, and cloning of fiber candidate genes. So far many QTLs for fiber quality have been identified and mapped. Details about this research were introduced in another chapter of this book. The cloning of some fiber candidate genes has been outlined in the previous section and elsewhere in this book. However, there has been little information about positional correspondence between the mapped genes/QTLs, and cloned fiber related cDNA/genes (fiber candidate genes). Recently, we have done a meta-analysis of fiber related QTLs and mutants, and compared their positions with those of cloned fiber candidate genes utilizing synteny between cotton and *Arabidopsis* (Rong et al., 2007). A total of 224 fiber-related QTLs and 7 fiber mutant genes were plotted on a high density reference map (Rong et al., 2004) and further to a consensus map (Rong et al., 2005a) of a hypothetical common ancestor of A and D genome cottons. In parallel, a total of 203 *Arabidopsis* homologs, representing up to the four best matches (at $E < 10^{-10}$) for each of 78 fiber candidate genes known from prior published work to be preferentially expressed during cotton lint fiber growth and development,

and 28 *Arabidopsis* genes known to be critical for trichome development, were determined. These *Arabidopsis* homologs were plotted on *Arabidopsis* alpha duplicated segments mentioned in section 4 and the established syntenic relationship between cotton and *Arabidopsis* (Rong et al., 2005a) was used to investigate their correspondence to the mapped fiber-related QTLs. Many such correspondences were identified and can be found in a CMap database (<http://www.plantgenome.uga.edu/cmap>). We found a general association between concentrations of candidate genes, and concentrations of cotton fiber-related QTLs. Based on synteny inferred using CrimestatII (Rong et al., 2005a), a significant correlation ($r = 0.260$, $N=167$, $p = 0.0003455$), was found between QTL number in cotton and candidate gene number in *Arabidopsis*. When we considered only the smaller number (largely a subset) of regions of synteny inferred using FISH (Rong et al., 2005a), the correlation was smaller ($r = 0.072$, $N=189$) and fell short of significance ($p=0.164$). In both CS2 and FISH-based models, the correlations between numbers of fiber QTLs and all (not just candidate) *Arabidopsis* genes in the syntenic regions detected by CS2 or FISH were not significant, suggesting that the relationship is specific to the population of candidate genes.

6.3 Broader Applications – Identification and Isolation of Interesting Genes for Cotton Breeding

Arabidopsis as a model plant has been studied extensively and the molecular mechanisms for many metabolite pathways are growing clearer, such as salt tolerance (Tester and Davenport, 2003), disease resistance (Cunnac et al., 2007) and floral development (Running and Hake, 2001) which are very important characters for cotton fiber production. Comparative messages between *Arabidopsis* and crops such as cotton will be useful in identification and isolation of these agriculturally important genes for crop breeding.

Cotton, in comparison to other crop plants, is classified as a relatively salt-tolerant crop, but not a halophyte. Cotton is native to subtropical or tropical zones where rainfall is normally deficient during the cotton growing season and plants can regularly be subjected to natural selection by drought and high salt content in some areas. Understanding the genetic basis of drought and salt tolerance has practical significance in cotton production. Although *Arabidopsis* is a typical glycophyte that is not particularly salt tolerant, recent studies suggest that it may contain most, if not all, of the salt tolerance genes that one might find in halophytes (Taji et al., 2004). Salt tolerance mechanisms in *Arabidopsis* have been explored and several related genes have been characterized (Tester and Davenport, 2003; Ward et al., 2003). Dehydration responsive element binding proteins (DBPs) are members of a larger family of transcription factors that are specific to plants and play an important role in enhancing plant tolerance to environmental stresses such as drought, cold, and high

salinity (Liu et al., 1998). With the combination of bioinformatic method and RACE-PCR technique, one cDNA encoding DRE-binding transcription factor, designated GhDBP3, was cloned from *G. hirsutum* with the help of DREB1A amino acid sequences (AB007787), a DRE-binding transcription factor of *Arabidopsis* (Huang and Liu, 2006). Further study indicated that GhDBP3 may play an important role in response to ABA and environment stress (Huang and Liu, 2006). This is one example of how knowledge of *Arabidopsis* molecular biology may expedite the isolation of cotton homologous genes.

Because the *Arabidopsis* genome is fully sequenced, cotton-*Arabidopsis* synteny will be useful in enriching the map of a target region of the cotton genome by identification of cotton ESTs homologous to *Arabidopsis* genes in the region. This strategy has been adopted in finding the markers close to the fiber mutant gene *Lil* mapped on the middle region of Chr.22 (Rong et al., 2005b). The *Lil* region showed synteny with *Arabidopsis* alpha 11 and 14 duplications. *Arabidopsis* genes on the syntenic regions were blasted to cotton ESTs, and the best EST hits ($E < 10^{-10}$, matching length > 100 bps) were used as probes to hybridize to cotton BAC libraries. The synteny-selected ESTs co-hybridized to BACs that also contained genetically-mapped markers, significantly more frequently than did randomly selected cotton ESTs, suggesting that cotton-*Arabidopsis* synteny is useful to enrich a targetted region for DNA markers (Rong and Paterson, unpublished data).

In summary, improved knowledge of the evolutionary histories of both *Arabidopsis* and cotton, has improved our ability to make deductions about the arrangement of corresponding genes in these two genomes. This information aids cotton researchers in taking better advantage of extensive functional information for *Arabidopsis* genes, and in accelerating isolation of important cotton genes. The eventual sequencing of one or more cotton genomes will further aid such translational genomic studies. There is considerable potential in revealing the genetic basis of conserved biological pathways using comparative genomics of cotton and *Arabidopsis*. The detailed comparisons that will be made possible by an eventual completed cotton genome may also permit us to identify cotton genes that have diverged particularly greatly, or in specific amino acids or regulatory motifs that imply striking changes to function. Such genes may warrant further testing to explore their influence on unique features of cotton that differentiate it from other plants and contribute to its economic importance.

References

- Nikolajeva, A. (1923). A hybrid between Asiatic and American cotton plant *Gossypium herbaceum* L. and *Gossypium hirsutum* L. *Bull. Appl. Bot. Plant Breed.* 13, 117–134.
- Denham, H. J. (1924). The cytology of the cotton plant. I. Microscopore formation in Sea Island cotton. *Ann. Bot. (London)* 38, 407–432.

- Longley, A. E. (1933). Chromosomes of *Gossypium* and related genera. *J. Agric. Res. (Washington, D.C.)* 46, 217–227.
- Skovsted, A. (1933). Cytological studies in cotton. I. The mitosis and meiosis in diploid and triploid Asiatic cotton. *Ann. Bot. (London)* 2847, 227–251.
- Skovsted, A. (1934). Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet.* 28, 407–424.
- Webber, J. M. (1935). Interspecific hybridization in *Gossypium* and the behavior of F1 plants. *J. Agric. Res. (Washington, D. C.)* 51, 1047–1070.
- Skovsted, A. (1937). Cytological studies in cotton. IV. Chromosome conjugation in interspecific hybrids. *J. Genet.* 34, 97–134.
- Beasley, J. O. (1940). The origin of American tetraploid *Gossypium* soecies. *Am. Nat.* 74, 285–286.
- Harland, S. C. (1940). New allopolyploids in cotton by the use of colchicine. *Trop. Agric.* 17, 53–55.
- Fryxell, P. A. (1963). Morphology of the base of seed hairs of *Gossypium*. I. Cross morphology. *Bot. Gaz.* 123, 196–199.
- Stebbins, G. (1966). Chromosomal variation and evolution; polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. *Science* 152, 1463–1469.
- Johnson, H. B. (1975). Plant pubescence and ecological perspective. *Bot. Rev.* 41, 233–258.
- Gottlieb, L. D. (1982). Conservation and duplication of isozymes in plants. *Science* 216, 373–380.
- Endrizzi, J. E., Turcotte, E. L., and J., K. R. (1985). Genetics, cytogenetics, and evolution of *Gossypium*. *Adv. Genet.* 23.
- Wendel, J. F. (1989). New World Tetraploid Cottons Contain Old-World Cytoplasm. *Proc. Natl. Acad. Sci. USA* 86, 4132–4136.
- Wendel, J., and Percival, A. E. (1990). Molecular divergence in the Galapagos Island-Baja California species pair, *Gossypium klotzschianum* Anderss., and *G. davidsonii* Kell. *Plant Syst. Evol.* 171, 99–115.
- Arumuganathan, K., and Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant. Mol. Biol. Rep.*, 208–218.
- McGrath, J. M., Jancso, M. M., and Pichersky, E. (1993). Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis-thaliana*. *Theor. Appl. Genet.* 86, 880–888.
- Kishimoto, N., Higo, H., Abe, K., Arai, S., A. Saito, A., and K. Higo, K. (1994). Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. *Theor. Appl. Genet.* 88, 722–726.
- Kowalski, S., Lan, T.-H., Feldmann, K., and Paterson, A. H. (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. *Genetics* 138, 499–510.
- Reinisch, A., Dong, J.-M., Brubaker, C., Stelly, D., Wendel, J., and Paterson, A. (1994). A detailed RFLP map of cotton (*Gossypium hirsutum* x *G. barbadense*): Chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138, 829–847.
- Wendel, J. F., Schnabel, A., and Seelanan, T. (1995). An Unusual Ribosomal DNA Sequence from *Gossypium gossypioides* Reveals Ancient, Cryptic, Intergenomic Introgression. *Mol. Phylogenet. and Evol.* 4, 298–313.
- Gaut, B. S., Morton, B. R., McCaig, B. C., and Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc. Natl. Acad. Sci. USA* 93, 10274–10279.
- Paterson, A. H., Lan, T. H., Reischmann, K. P., Chang, C., Lin, Y. R., Liu, S. C., Burow, M. D., Kowalski, S. P., Katsar, C. S., DelMonte, T. A., Feldmann, K. A., Schertz, K. F., and Wendel, J. F. (1996). Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics* 14, 380–382.

- Marks, M. D. (1997). Molecular genetic analysis of trichome development in arabidopsis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48, 137–163.
- Mauricio, R., and Rausher, M. D. (1997). Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defence. *Evolution* 51, 1435–1444.
- Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* 10, 1391–1406.
- Muravenko, O. V., Fedotov, A. R., Punina, E. O., Fedorova, L. I., Grif, V. G., and Zelenin, A. V. (1998). Comparison of chromosome BrdU-Hoechst-Giemsa banding patterns of the A1 and (AD) (2) genomes of cotton. *Genome* 41, 616–625.
- Brubaker, C. L., Paterson, A. H., and Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42, 184–203.
- Hsu, C. Y., Creech, R. G., Jenkins, J. N., and Ma, D. P. (1999). Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. *Plant Science* 143, 63–70.
- Hulskamp, M., Schnittger, A., and Folkers, U. (1999). Pattern formation and cell differentiation: Trichomes in Arabidopsis as a genetic model system. In “International Review of Cytology – a Survey of Cell Biology, Vol 186”, Vol. 186, pp. 147–178.
- Loguerccio, L. L., Zhang, J. Q., and Wilkins, T. A. (1999). Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). *Mol. Gen. Genet.* 261, 660–671.
- Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., Esch, J. J., Marks, M. D., and Gray, J. C. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1349.
- Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498.
- Liu, H. C., Creech, R. G., Jenkins, J. N., and Ma, D. P. (2000). Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* 1487, 106–111.
- Paterson, A. H., Bowers, J., Burow, M., Draye, X., Elsik, C., Jiang, C., Katsar, C., Lan, T., Lin, Y., Ming, R., and Wright, R. (2000). Comparative genomics of plant chromosomes. *Plant Cell* 12, 1523–1539.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Vision, T., Brown, D., and Tanksley, S. (2000). The origins of genomic duplications in Arabidopsis. *Science* 290, 2114–2117.
- Eckardt, N. A. (2001). Everything in its place: Conservation of gene order among distantly related plant species. *Plant Cell* 13, 723–725.
- Running, M. P., and Hake, S. (2001). The role of floral meristems in patterning. *Curr. Opin. Plant Biol.* 4, 69–74.
- Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R. L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J. P., Miguel, T., Paszkowski, U., Zhang, S. P., Colbert, M., Sun, W. L., Chen, L. L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y. S., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). *Science* 296, 92–100.

- Levine, N. (2002). CrimeStat: A spatial statistics program for the analysis of crime incident locations (v 2.0). *Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC*.
- Oh, K., Hardeman, K., Ivanchenko, M., Ellard-Ivey, M., Nebenfuhr, A., White, T., and Lomax, T. (2002). Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the Diageotropica (Dgt) locus. *Genome Biology* 3, research0049.1 – research0049.11.
- Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jurgens, G., and Hulskamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *Embo Journal* 21, 5036–5046.
- Vandepoele, K., Saeys, Y., Simillion, C., Raes, J., and Van de Peer, Y. (2002). The Automatic Detection of Homologous Regions (ADHoRe) and Its Application to Microcolinearity Between *Arabidopsis* and Rice. *Genome Res.* 12, 1792–1801.
- Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., Marks, M. D., Shimura, Y., and Okada, K. (2002). Role of a positive regulator of root hair development, CAPRICE, in *Arabidopsis* root epidermal cell differentiation. *Development* 129, 5409–5419.
- Yu, J., Hu, S. N., Wang, J., Wong, G. K. S., Li, S. G., Liu, B., Deng, Y. J., Dai, L., Zhou, Y., Zhang, X. Q., Cao, M. L., Liu, J., Sun, J. D., Tang, J. B., Chen, Y. J., Huang, X. B., Lin, W., Ye, C., Tong, W., Cong, L. J., Geng, J. N., Han, Y. J., Li, L., Li, W., Hu, G. Q., Huang, X. G., Li, W. J., Li, J., Liu, Z. W., Liu, J. P., Qi, Q. H., Liu, J. S., Li, T., Wang, X. G., Lu, H., Wu, T. T., Zhu, M., Ni, P. X., Han, H., Dong, W., Ren, X. Y., Feng, X. L., Cui, P., Li, X. R., Wang, H., Xu, X., Zhai, W. X., Xu, Z., Zhang, J. S., He, S. J., Zhang, J. G., Xu, J. C., Zhang, K. L., Zheng, X. W., Dong, J. H., Zeng, W. Y., Tao, L., Ye, J., Tan, J., Ren, X. D., Chen, X. W., He, J., Liu, D. F., Tian, W., Tian, C. G., Xia, H. G., Bao, Q. Y., Li, G., Gao, H., Cao, T., Zhao, W. M., Li, P., Chen, W., Wang, X. D., Zhang, Y., Hu, J. F., Liu, S., Yang, J., Zhang, G. Y., Xiong, Y. Q., Li, Z. J., Mao, L., Zhou, C. S., Zhu, Z., Chen, R. S., Hao, B. L., Zheng, W. M., Chen, S. Y., Guo, W., Li, G. J., Liu, S. Q., Tao, M., Zhu, L. H., Yuan, L. P., and Yang, H. M. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). *Science* 296, 79–92.
- Blanc, G., Hokamp, K., and Wolfe, K. H. (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* 13, 137–144.
- Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438.
- Calabrese, P. P., Chakravarty, S., and Vision, T. J. (2003). Fast identification and statistical evaluation of segmental homologies in comparative maps. *Bioinformatics* 19 Suppl. 1, i74–i80.
- Cannon, S., Kozik, A., Chan, B., Micheltore, R., and Young, N. (2003). DiagHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization. *Genome Biology* 4, R68.
- Larkin, J. C., Brown, M. L., and Schiefelbein, J. (2003). How do cells know what they want to be when they grow up? Lessons from Epidermal Patterning in *Arabidopsis*. *Annu. Rev. Plant. Biol.* 54, 403–430.
- Paterson, A., Bowers, J., Peterson, D., Estill, J., and Chapman, B. (2003). Structure and evolution of cereal genomes. *Curr. Opin. Genet. Dev.* 13, 644–650.
- Schiefelbein, J. (2003). Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. *Curr. Opin. Plant Biol.* 6, 74–78.
- Suo, J. F., Liang, X., Pub, L., Zhang, Y. S., and Xue, Y. B. (2003). Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* 1630, 25–34.
- Tester, M., and Davenport, R. (2003). Na⁺ tolerance and Na⁺ transport in higher plants. *Ann. Bot.* 91, 503–527.

- Ward, J. M., Hirschi, K. D., and Sze, H. (2003). Plants pass the salt. *Trends in Plant Sci.* 8, 200–201.
- Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78, 139–186.
- Blanc, G., and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678.
- Hulskamp, M. (2004). Plant trichomes: A model for cell differentiation. *Nature Rev. Mol. Cell Biol.* 5, 471–480.
- Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004a). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* 101, 9903–9908.
- Paterson, A. H., Bowers, J. E., Chapman, B. A., Peterson, D. G., Rong, J. K., and Wicker, T. M. (2004b). Comparative genome analysis of monocots and dicots, toward characterization of angiosperm diversity. *Curr. Opin. Biotechnol.* 15, 120–125.
- Rong, J., Abbey, C., Bowers, J. E., Brubaker, C. L., Chang, C., Chee, P. W., Delmonte, T. A., Ding, X. L., Garza, J. J., Marler, B. S., Park, C.-H., Pierce, G. J., Rainey, K. M., Rastogi, V. K., Schulze, S. R., Trolinder, N. L., Wendel, J. F., Wilkins, T. A., Williams-Coplin, T. D., Wing, R. A., Wright, R. J., Zhao, X., Zhu, L., and Paterson, A. H. (2004). A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166, 389–417.
- Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J.-K., and Shinozaki, K. (2004). Comparative Genomics in Salt Tolerance between *Arabidopsis* and *Arabidopsis*-Related Halophyte Salt Cress Using *Arabidopsis* Microarray. *Plant Physiol.* 135, 1697–1709.
- Wang, S., Wang, J.-W., Yu, N., Li, C.-H., Luo, B., Gou, J.-Y., Wang, L.-J., and Chen, X.-Y. (2004). Control of Plant Trichome Development by a Cotton Fiber MYB Gene. *Plant Cell* 16, 2323–2334.
- Hsu, C.-Y., Jenkins, J. N., Saha, S., and Ma, D.-P. (2005). Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein. *Plant Science* 168, 167–181.
- Humphries, J. A., Walker, A. R., Timmis, J. N., and Orford, S. J. (2005). Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* TRANSPARENT TESTA GLABRA1 (TTG1) gene. *Plant Mol. Biol.* 57, 67–81.
- Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70.
- Rong, J., Bowers, J. E., Schulze, S. R., Waghmare, V. N., Rogers, C. J., Pierce, G. J., Zhang, H., Estill, J. C., and Paterson, A. H. (2005a). Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Res.* 15, 1198–1210.
- Rong, J., Pierce, G., Waghmare, V., Rogers, C., Desai, A., Chee, P., May, O., Gannaway, J., Wendel, J., Wilkins, T., and Paterson, A. (2005b). Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. *Theor. Appl. Genet.* 111, 1137–1146.
- Huang, B., and Liu, J.-Y. (2006). Cloning and functional analysis of the novel gene GhDBP3 encoding a DRE-binding transcription factor from *Gossypium hirsutum*. *Biochim. Biophys. Acta, Gene Struct. Expr.* 1759, 263–269.
- Odland, W., Baumgarten, A., and Phillips, R. (2006). Ancestral rice blocks define multiple related regions in the maize genome. *Crop Sci* 46, S-41–48.
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A.,

- Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E., Unneberg, P., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A. B., Sickler, B. A., Wilkins, T. A., Guo, J. Y., Chen, X. Y., Scheffler, J., Taliencio, E., Turley, R., McFadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S. R., Pierce, M. L., Essenberg, M., Kim, H., Llewellyn, D. J., Dennis, E. S., Kudrna, D., Wing, R., Paterson, A. H., Soderlund, C., and Wendel, J. F. (2006). A global assembly of cotton ESTs. *Genome Res.*, gr. 4602906.
- Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W., Ge, S., and Luo, J. (2006). Statistical inference of chromosomal homology based on gene colinearity and applications to *Arabidopsis* and rice. *BMC Bioinformatics* 7, 447.
- Cunnac, S., Wilson, A., Nuwer, J., Kirik, A., Baranage, G., and Mudgett, M. B. (2007). A conserved carboxylesterase is a suppressor of avrbst-elicited resistance in *Arabidopsis*. *Plant Cell* 19, 688–705.
- Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., and Paterson, A. H. (2007). Meta-analysis of Polyploid Cotton QTLs Shows Unequal Contributions of Subgenomes to a Complex Network of Genes and Gene Clusters Implicated in Lint Fiber Development. *Genetics* doi: 10.1534/genetics.107.074518.

Impacts on Agroecosystems of Transgenic Insect and Herbicide Resistance in Cotton

Lawrence C. Davis

Abstract Cotton is the world's major source of natural fiber. Enhanced production will require both a full understanding of the genomic potential of cotton species, and deployment of transgenes for specific traits. Genomic approaches documented elsewhere in this volume will allow identification of key genes in the process of capturing and converting photosynthetic energy to usable fiber. Genomic approaches also provide tools to target specific genes for altered regulation. As a first step to understanding the complex functions of a plant, introduction of transgenes can be highly effective and informative. Cotton, which requires intensive management for high yield production, is one of the few crops for which insect and herbicide resistance traits have been commercialized. Benefits to producers in highly mechanized agricultural systems have been substantial and adoption is widespread. Advantages in smaller scale production systems are much less clear. Because there are huge expenditures required to get transgenic plants into widespread commercial production, it is instructive to carefully examine the impacts of the currently available examples. Specific issues may differ but economic and sociopolitical factors will require similar consideration, whatever the trait being modified. Concerns have been expressed that there may be direct impacts of transgenes, such as insect or herbicide resistance genes, on non-target organisms, damage to insect predators, or risks of gene flow to wild species. Examples from around the world, in many different climates and agricultural systems have been reviewed. No scientifically solid evidence has been forthcoming to indicate that the gene constructs or cultivars per se have a negative impact on agroecosystems, including non-target arthropods, predators or wild cotton species. Heavy herbicide use associated with herbicide-resistant cotton may lead to increased populations of resistant weeds, but no-till production systems based on herbicide use may provide large benefits. Other interesting traits including disease resistance have been introduced to cotton but none have yet been commercialized or extensively tested at a field scale.

L.C. Davis (✉)

Biochemistry Department, Kansas State University, Manhattan, KS 66506
e-mail: ldavis@ksu.edu

1 Cotton Productivity Potential

Cotton is the world's largest source of natural fiber. To meet the rising demand for clothing and textiles in developing economies, production must increase, or substitution by other (primarily chemically synthesized) fibers will occur. Such substitution is very sensitive to relative price, so that variations in yield and input costs for cotton, versus the price of petroleum feedstocks, can lead to large swings of supply and demand. With a relatively fixed available area, on the order of 75 million acres (30 million hectares) worldwide, increased yield is the key to cotton maintaining its market share (Cantrell 2006). Yields worldwide vary enormously, with Australia posting values of over 1800 kg/ha, Pakistan having about 650 kg/ha and India lagging at below 500 kg/ha until very recently. (Ahmad 2007). In the U.S. there is similar wide variation from one state to another with an average of around 800 lb/A (~880 kg/ha). The world average is about 700 kg/ha, indicating a good potential for increased yields. China, the world's largest consumer of cotton, with a large domestic production, has yields of over 1100 kg/ha. Many producers in Africa, Asia and South America are below 500 kg/ha. Most of the above yield statistics are recent averages derived from published data (Johnson, MacDonald, McKeivier, Meyer and Skelly 2006) and from USDA Foreign Agricultural Service online reports.

As with other major crops, rates of increase in cotton fiber yield through use of conventional breeding and production technologies will reach a plateau. If that plateau were simply a function of photosynthate production and storage, a yield of 6000 kg/ha (e.g. as found with sugar from beets) would not be unreasonable under good growing conditions. In other C-3 crops such as irrigated alfalfa, full season biomass yield regularly exceeds 30,000 kg/ha in climates where cotton is grown, while cotton biomass is about 1/3 as much (Sadras and Wilson 1997). At present the cotton seed yield is about 1.6 times the fiber yield by weight, with an overall harvest index of ~0.3 for fiber plus seed. Clearly there is still much room for improvement in partitioning of photosynthate to the desired product, and we have limited understanding of factors that control partitioning to reproductive structures, or to fiber versus seed. Application of new genomic strategies including identification of regulatory elements, selective modification of genes, or introduction of novel genes, will be needed to advance beyond the present crop-specific barriers.

Other chapters in this volume detail efforts to determine the full genomic potential of the cotton species currently being used for commercial production, and as exotic germplasm sources. Here we focus on a few examples of how introduction of DNA from widely distant species may have major impacts on cotton production. Thus far, only insect resistance and herbicide resistance traits have been commercialized. As discussed in detail below, these have been effective and also relatively well accepted by both producers and consumers.

Thus, other modifications of the cotton plant as a producer of cotton fiber may also prove acceptable. There may be additional concerns when cottonseed meal or oil is used in animal feed or human food.

2 Costs Associated with Cotton Production

Major expenditures in cotton production include insecticides and herbicides. Compared to the food crops wheat, corn and soybeans, the cost per area (and presumed quantity used) for insecticides and herbicides is several-fold higher in cotton (USDA ERS 2004). Historically, cotton had a reputation of requiring heavy and frequent pesticide and herbicide applications. For instance, in the late 1980s, a typical U.S. farm might use seven pesticide and three to five herbicide applications, even with some integrated pest management strategies in place (Frisbie, El-Zik and Wilson 1989). On the other hand, it was also observed that for Texas, only 1/3 of the cotton area received pesticides in the early 1980s, and in California the average was 1 or 2 sprayings per year (Frisbie et al 1989). Recently in India, pesticides were applied up to 30 times per season in some districts, and may have represented half of production costs (Shetty 2004). Some of the pesticides in use world-wide are banned in the U.S., on the basis of their health and environmental hazards (Burcher 2007). These represent societal costs, rather than direct producer costs.

Application of materials to crops naturally has associated fuel and labor costs. Thus, the number of applications, as well as the amounts applied, affects production costs. According to the USDA Economic Research Service (Shoemaker, McGranahan and McBride 2006) direct energy costs for cotton were over 3.5 times higher than for soybeans (\$64 vs \$18/acre, or \$150 vs \$40/ha) in regions where both are grown. Such direct energy costs accounted for only 1/5 of the total cost of either cotton or soybean production, but cotton yields a higher dollar value per area of crop than soybeans, while requiring larger energy inputs. Direct energy costs are particularly sensitive to the price of oil, but fertilizer and irrigation are also highly energy dependent.

3 Potential for Benefits from Biotech Cotton

Given the high costs of production and the potentially large yield increases of cotton that could be obtained by control of pests and weeds, without repeated applications of pesticide and herbicide, the introduction of transgenes for insect and herbicide resistance appeared to be a logical thing to do for cotton when the euphoria for biotech crops was at its peak during the 1980s and 1990s (Jenkins 1999). It was anticipated that insect resistance could dramatically decrease the need for insecticides by making the cotton plant resistant to some of its major pests, such as the bollworm, budworm and armyworm. Making the plant

resistant to certain broad-spectrum herbicides could allow better control of a wide range of weeds with fewer treatments. It would also promote minimum tillage production systems, whereby herbicide would substitute for tilling in weed management. This could reduce direct energy costs dramatically and reduce soil compaction and erosion. Because cotton uses a large fraction of crop land in some areas, and often is grown on erodible soils, these changes would have a large impact in some agroecosystems, reducing soil erosion and perhaps increasing insect biodiversity.

High quality cotton can only be produced with good attention to detail, including factors such as the length, color, and stickiness of the fibers. These are sensitive to insects and stresses. Thus cotton has a reputation of requiring relatively high doses of insecticides, and low weed competition for maximum yield. Stickiness afflicts a significant portion (20-25 %) of world cotton production (Hequet and Abidi 2006) and is a result of exudates by whitefly and aphid populations. Other pests may cause stains, affecting the color of the cotton. It may be speculated that if Bt reduces insecticide use, some secondary pests may also be reduced, because excess insecticide use is the likely cause of aphid and whitefly population explosions (Long, Flint, Layton and Steinkraus 2002). As an alternative to insecticides, or enhancement of predator populations, both aphids and whiteflies may be controlled by entomopathogenic fungi (MS State U 2007a; NCSU 1999).

These two examples of transgenic cotton were initially developed early in the larger development of transgenic methods. More recent advances in understanding of genomes have come from the many DNA sequencing efforts that have been applied to major crop plants. We now have a much better understanding of metabolic pathways at the molecular level, including roles of the Cesa and other auxiliary genes that are involved in cellulose fiber production (Saxena and Brown 2005). It is anticipated that modification of such pathways holds the potential for increased cotton fiber production. However, going from concept to product has not been straightforward even for such "simple" traits as insect or herbicide resistance. The multiplicity of similar genes being identified in the Ces gene family, and the number of other factors identified as regulatory for cotton fiber production in cultures, indicates that a fuller understanding at the genomic level, including comparative genomics with model systems such as *Arabidopsis*, is essential.

4 Development and Adoption of Biotech Cotton

The adoption of biotech cotton cultivars (Cvs) has been fairly fast and quite extensive. Bollgard cotton, expressing *Bacillus thuringiensis* (Bt) Cry1 Ac delta endotoxin protein was commercially introduced in the U.S. in 1996, and in other countries in rapid succession. The expressed protein is selectively toxic to lepidopterans, particularly those that are major pests of cotton and maize. Over

1/2 of all world cotton production is now from biotech Cvs, although less than 1/2 of the total cotton producing area is in countries that permit use of such crops. In the U.S. it is estimated that biotech cotton was found on more than 80 % of all acreage in 2005 (Cantrell 2006). For 2006 over 95% of Upland cotton was transgenic (not only Bt, but including herbicide resistance), and only California (at 59%) had less than 80% transgenic plants among U.S. states (Cotton Council International 2007). Herbicide resistance was introduced, and has spread, over a similar time-frame to that of Bt (USDA AMS 2006).

Cantrell (2006) makes a useful distinction between innovation in inputs vs outputs. Thus far, only resistances to insects and herbicides have been commercialized in cotton. These represent production inputs and were obtained by introducing genes from other (bacterial) species. Resistance to fungi is another input for which some genetic modification steps have been taken (Wang, Chen, Wang, Huang, Yao, Liu, Wei, Li, Zhang, and Sun 2004). Alteration of gossypol levels, oil quality, fiber structure and other features of the harvested crop would represent outputs. Genomic methods of gene discovery are rapidly giving insight into the multiplicity of alleles and modes of regulation of cotton genes. Other chapters in this volume lay the basis for modification of such traits. Although there are numerous scientific publications, there are no commercialized products yet.

5 Biotech, but not Transgenic, Crops

There may be potential for modification of herbicide or other regulatory molecule targets within the cotton genome, as has been successfully done with soybean, maize, sunflower and wheat, without resort to genes from other species. In these crops, the enzyme acetolactate synthase has been modified by selection of natural or induced mutations. Clearfield technology (BASF 2007) is based on this principle, whereby imidazolinine or sulfonyleurea type herbicides are used for control of weeds in a crop carrying an enzyme that fails to bind the herbicide. This strategy of selectively altering an enzyme normally present in the crop plant by “conventional” mutagenesis and selection, avoids the U.S. testing and registration requirements because no transgenes are involved. Such crops still require registration in Canada as biotech crops, although they are not transgenic. One can use a Canadian government website (Canada 2006) to determine what crops have been modified with the intention of commercialization there. The contribution by Wright in this volume addresses the possibility of altering insect resistance in a similar fashion, by altering cotton genes *in vivo* or *in vitro*.

At this point (2008) biotech cotton without transgenes is not yet available in a commercial system, but recent reports of RNAi control of gossypol in seeds (Sunilkumar, Campbell, Puckhaber, Stipanovic and Rathore 2006) suggest that such crops could soon appear. Whether they are considered transgenic will

depend on how the modifications are produced and introduced to the plants, and how they are expressed. This point is not trivial, given sociopolitical realities.

6 Transgenic Cotton

This chapter examining published works, will be restricted to transgenes. Because of space limitations, preference is given to more recent publications which in turn provide access to earlier literature. There are three main features to be considered. Each may have its own ecological impacts. Insect resistance most obviously and immediately affects insects that feed on cotton. Herbicide resistance most immediately affects weeds that are competitive with cotton. However, in each case there are indirect as well as direct impacts of the transgene systems. Engineered disease resistance seems to directly affect only cotton, but there would be indirect effects on organisms that might take advantage of diseased cotton.

Among the obvious direct effects, for insect resistant Bt cotton, there is a decrease in the population of susceptible insects. This in turn results in a reduced food source for predators upon such insects. An indirect effect would be that by reducing the number of insecticide applications, the Bt cotton may have an increased population of other insects, pests or not. These in turn might provide a food source for predators. The predators for the two cases (Bt susceptible and unaffected insects) may be different insects, small mammals, or birds, because altering the relative population of Bt susceptible insects may alter populations of a number of other organisms within the same food web. Some of these potential ecological impacts were clearly recognized prior to introduction of Bt cotton (Fitt, Mares, and Llewellyn 1994).

A synthesized scorpion insect toxin AaHIT has been engineered into cotton to kill bollworm larvae (Wu, Luo, Wang, Tian, Liang, and Sun 2008). It is suggested as an alternative to Bt, but no information on the range of insects affected is given; only the bollworm was tested with leaves of transgenic plants. Similarly, a patent from Bayer Biosciences (van Rie 2006) reports that a bacterial insecticidal protein from *Xenorhabdus* in transgenic cotton is effective against *Helicoverpa zea*. For control of a secondary cotton pest, *Aphis gossypii*, a transgenic agglutinin from an amaranth is effective (Wu, Luo, Guo, Xiao, and Tian 2006). This transgene is meant to address the problem of sticky cotton. No field results are yet available for any of these three traits.

Herbicide resistance allows use of broad-spectrum herbicides such as Round-up (glyphosate) and Liberty (glufosinate). These herbicides kill a much broader spectrum of plants than do many older herbicides such as 2,4-D, atrazine or metolachlor. The most obvious impact in this instance is a reduction of biodiversity. The direct effect is to decrease populations of many species of plants. This in turn has the effect of reducing food sources for various

herbivores, including insects, small mammals and birds. A more insidious effect is that the few kinds of plants which have some natural or selected herbicide resistance may become much more abundant. Examples of such resistance include Palmer amaranth in Georgia, USA, and horseweed (*Conyza*) in Mississippi, USA, both having resistance to glyphosate. A current list, the international survey of herbicide-resistant weeds, may be viewed at the Weed Science Society of America website (<http://www.weedscience.org/in.asp>). New weeds are added frequently.

In most situations, economic benefits to producers are more influential than potential alterations of agroecosystems in determining whether a new technology is adopted and sustained. A number of studies have been directed at assessing relative benefits of different cropping strategies, including transgenic cotton (Raney 2006). Relatively few publications that measure ecological effects directly are available. Those studies are discussed below.

7 Resistance to Insects

7.1 Broad Effects of Bt Cotton

Thus far no insect resistance systems other than Bt have been commercialized in cotton, except in China where cowpea trypsin inhibitor (CpTI) is used to some extent (Wu and Guo 2005). The general term Bt is used to describe various genes of similar function derived from *Bacillus thuringiensis* species. The subscripts, as in Cry2 Ab or Cry1 Ac indicate similar but not identical proteins. Details of constructs actually used may be found on websites at either Agbios.com or Canada (2006). Most events used to develop resistant plants have contained one or more types of a crystalline delta endotoxin found in spores of the bacterium. Monsanto's Bollgard I has the Cry1 Ac while Bollgard II also contains Cry2 Ab. Widestrike (Dow Agrosiences) has Cry1 Ac plus Cry1 F. Both Bollgard II and Widestrike are designed to affect a broader range of lepidopterans than Bollgard. One transgenic event is listed that contains a vegetative protein designated Vip 3A (vegetative insecticidal protein), which works in a similar fashion as the delta endotoxin but binding to different sites in the lepidopteran gut (Lee, Walters, Hart, Palekar and Chen 2003). Syngenta is the developer of this type of cotton which is presently registered in the U.S. and Australia. Few publications have reported effects of Vip 3A cotton (Whitehouse, Wilson and Constable 2007). More information is available through registration documents filed with the USDA. Widestrike came on the market recently (Dow 2007) and was not used for any of the studies cited below.

Although there are Bt endotoxins that target non-lepidopterans, those used for cotton are believed to be specific. Nevertheless, some studies have been done to verify that the toxins, as expressed in cotton plants, do not significantly, directly affect populations of non-target insects or other arthropods. For

instance, Yu, Berry and Croft (1997) observed no effect of the Cry1 Ab gene product on a collembolan and an oribatid mite fed on residues of transgenic cotton. Torres and Ruberson (2008) showed that although detectable amounts of Cry1 Ab protein move up the trophic chain from plant, to lepidopteran larvae, to predatory soldier bugs (*Podisus maculiventris*), there is no effect on the life cycle of the predator. Indeed, in the absence of wide spectrum insecticides, predators may act synergistically with the Bt toxin to reduce lepidopteran larval populations.

One must also consider whether the alteration of plant composition may alter microbial flora, either prior to harvest, or as a function of residue incorporation into soil. One such report is available (Donegan, Palm, Fieland, Porteous, Ganio, Schaller, Bucaro, and Seidler 1995). Although two transgenic lines produced a transient increase in soil bacteria and fungi, another line and the purified delta endotoxin had no such effect. The authors (Donegan et al 1995) concluded that alterations of the host plant through cultural manipulations may have had an effect but that the transgene construct itself did not.

Rhizosphere effects of Bt toxin were directly addressed by Rui, Yi, Wang, Li, Zhai, He, and Li (2005). Some Bt toxin was detectable by ELISA in the root zone of Bt cottons, NuCOTN99 and SGK321 (derived from SY321). Microbial numbers were reduced in the rhizosphere of NuCOTN99 compared to SY321, but adding Bt toxin to rhizosphere soil directly did not produce the same alteration. The authors concluded that as seen by Donegan et al (1995) there was an effect of Cv, not of Bt gene per se.

While cotton is grown specifically for its fiber, a significant portion of the crop is not fiber but seed. The seed is of value for its polyunsaturated oil content and the residual cottonseed meal is used as a protein source where gossypol toxicity is not an issue. The Cry1 A gene is expressed in the cotton seed, hence cottonseed meal contains the toxin, although it could be in a denatured form if heat or solvent extraction is applied to maximize oil recovery. Potential toxicities must be considered when the several alternate toxins mentioned previously (from scorpion, *Xenorhabdus* and amaranth) are introduced to the cotton crop.

In China, where a large fraction of the cotton crop is transgenic, cottonseed meal is a major constituent of the growth substrate for edible fungi. One of the pests of such fungi, a dipteran *Lycoriella pleuroti* was found to be adversely affected by transgenic (Bt) cottonseed meal (Xu, Wang, Liu, Yuan, Su, Chen, Wu and Zhang 2006). The main effects were due, not to the toxin per se, but to altered levels of gossypol and condensed tannins, which are elevated in the transgenic line being studied. Growth medium containing cottonseed was more toxic than medium without it, and transgenic cottonseed meal was more toxic than that from the non-transgenic progenitor Cv. Purified Bt protein (Cry1 Ac) at the lowest tested level (0.1 mg/kg) produced similar toxicity compared to a 10-fold higher level, and the flavonoid quercetin also gave a non-linear dose response with maximum toxicity at the lowest dose of 0.1%. For gossypol and tannins, mortality increased with increasing content. For the

growers of the edible *Pleurotis* fungus this cottonseed toxicity provides an unintended benefit of the transgenic cottonseed (Xu et al 2006). However, it also indicates that the process of developing transgenic lines may significantly alter the cotton seed in ways not obviously related to the Bt genes. Such alteration of the plant is consistent with the report (Donegan et al 1995) where the microbial population on the residues was different for different derived lines.

7.2 Impact of Bt Cotton on Pests and Predators

Cotton cultivation occurs around the world in rainfed and irrigated conditions, under tropical, sub-tropical and temperate climates, in large-scale monoculture, intercropped, and as small plots. Regional examples are given below. Transgenic cottons may have different impacts depending on the other components of the ecosystem. Financial constraints have usually limited the intensity of sampling at a single site and the number of sites considered in a single study.

Recently, Marvier, McCreedy, Regetz and Kareiva (2007) have made a meta-analysis of available studies for both cotton and maize that meet specified criteria for inclusion in a database of Bt nontarget effect studies. That database is available at <http://delphi.nceas.ucsb.edu/btcrops/>. A total of 42 studies on cotton or maize were included in their analysis. Eight were for cotton, of which seven are discussed below. Only one cotton study conducted in Asia and one from Australia met their criteria. By measuring effect sizes for general classes of insects, under different treatment conditions, it was possible to derive statistical confidence limits for the impact of Bt cotton vs insecticide-treated cotton.

There was a small, but significant reduction in abundance of total non-target (i.e. mainly non-lepidopteran) invertebrates on Bt compared to non-Bt cotton without insecticide, no difference when both Cvs received insecticides, and an increase when only the non-Bt cotton received insecticide applications. When examined across orders of arthropods, there were significant reductions in Coleoptera and Hemiptera, when comparing Bt and non-Bt Cvs, both grown without insecticide treatment. These orders include non-target species (e.g. damselbugs and ladybird beetles) that may have lower abundance because they have fewer prey available. Non-target lepidopterans were markedly reduced also, as expected. Individual reports are considered below.

7.2.1 Southeastern, Rainfed U.S

The southeastern part of the U.S. is historically a very early, although not the original, site of U.S. cotton production, which was initially focused around Virginia. Cotton growing has had an impact on agroecosystems of the southeast for 200 years (Frisbie et al 1989). Predaceous arthropods were monitored under non-disrupted and disrupted conditions on Bollgard and Bollgard II cotton in comparison to the parental Cv by Hagerty, Kilpatrick, Turnipseed, Sullivan

and Bridges (2005). Bollgard II contains both CryI Ac, like Bollgard, plus Cry2 Ab which broadens its range of efficacy to fall armyworms (*Spodoptera*) and soybean loopers (*Pseudoplusia*) with enhanced effect against the bollworm, *Helicoverpa zea*, which has corn as an alternate host plant. Predator numbers were increased or unchanged on untreated Bollgard and Bollgard II plots in each of three seasons. When a broad spectrum insecticide was used early in the season to disrupt the predator population, numbers of the pest *H. zea* exceeded economic thresholds in one season with a high population, even on the Bollgard II. With such extreme infestation, predator and other pest populations declined on the parental line late in the season, giving a significantly lower total population of predators. The authors concluded that it is important to minimize use of broad spectrum insecticides to maximize the benefits of Bt toxins for control of lepidopterans.

Torres and Ruberson (2005) compared three pairs of commercial Bt and non-Bt fields in Georgia, USA in each of the years 2002-2004, monitoring abundance of both canopy- and ground-dwelling predators. All fields were > 5 ha in area. Samples were taken throughout the growing season, using three methods: bagging of whole plants, use of drop cloths, and pitfall traps. The sampling was intensive, with 40 drop cloth samples per week per field, 20 pit traps set as pairs along a 10 point transect across the field, and 20-30 whole plants per field bagged in plastic and taken near the 10 sampling points. From 18 to 30 taxa were collected by the three methods. Considering cumulative data for three whole seasons, two species showed a significant increase on the Bt cotton by one or another of the methods, while three species were more abundant on non-Bt cotton. For some species, or life stages, only one of the three methods was effective. For instance only whole plant inspection revealed egg masses of predators, and only pitfall traps yielded some ground-dwelling species. Because the fields were in commercial production, insecticide was used as needed, according to an IPM model, on the non-Bt cotton. Insecticide application selectively reduced some species and appeared to have a larger impact than the Bt cotton on species abundance. Overall, there was no evidence of a negative impact of Bt cotton on predator populations.

Looking over a wider geographic area (South Carolina, Georgia, north and south Alabama), Head, Moar, Eubanks, Freeman, Ruberson, Hagerty and Turnipseed (2005) compared whole arthropod populations on large commercial fields of Bt versus non-Bt cotton in years 2000-2002. The general design for these studies was similar to the work of Torres and Ruberson (2005) but with fewer sampling sites per field and sampling over a shorter portion of the growing season. Numbers of predators varied widely, over seasons and by locations. Fire ants, spiders and ladybeetles were generally the most abundant predators, with several kinds of predatory bugs also present in significant numbers. Fire ant numbers were directly related to number of insecticide applications, which was reduced from about 2.4 times per season on non-Bt to 1.1 per season on Bt cotton.

Target pests were much less abundant in Bt than non-Bt fields in 8 of 12 comparisons (3 seasons, 4 locations). For non-target pests, plant bugs (*Lygus*) were more abundant in Bt fields in 5 of 12 and less abundant in 4 of 12 comparisons, while stink bugs, as nymphs or adults, were more abundant in 2 or 3 of 12 comparisons. The Bt cotton received no insecticide even when the relative abundance of the plant bug adults was 15 times that in non-Bt cotton because the absolute numbers were small ($\sim 1/100$ m row for *Lygus* in non-Bt cotton and $\sim 1/50$ m row for stink bugs). Economic thresholds are in the range of 1-2/ m row, differing by state and stage of plant growth (MS State U 2007b).

Predation rate is an important measure of how Bt cotton may impact the agroecosystem. Disappearance of deliberately introduced lepidopteran eggs was greater in Bt than non-Bt fields, often by a very large factor (e.g. 90 % versus 50%). Thus, use of Bt cotton had a positive impact on insect control.

7.2.2 Arizona, Irrigated Cotton

Naranjo and Ellsworth (2002) addressed the impact of Bt cotton (Bollgard and Bollgard II) on arthropod communities in the western U.S. where cotton typically received a dozen applications of pesticides prior to introduction of Bt cotton, which required only about two. The study design included use of isogenic cotton (parent line) without Bt. During 1999 there was a trend toward lower predator populations in the Bt cotton but in 2000 it was reversed. Neither trend was significant and overall it amounted to about 10 % of total individuals, with somewhat greater or lesser differences by species. Total predator abundance was nearly three-fold greater in 1999 than in 2000. The two predator species with lower populations are large predators that prey on caterpillars and may have found insufficient food to remain in the experimental plots.

Total species richness and diversity did not differ between years, but insecticide use did reduce “richness” (species number) in one year and increase “evenness” in both years, by decreasing the numbers of individuals for the most abundant species. Rates of predation on pests did not differ between Bt and non-Bt cotton in 2001, and no significant parasitism was observed on sentinel egg placements in either instance. Naranjo and Ellsworth (2002) suggest that the Bt effect is highly selective, not altering overall community structure.

Carriere, Ellers-Kirk, Sisterton, Antilla, Whitlow, Dennehy, and Tabashnik (2003), did a long-term study, comparing five years after introduction of Bt cotton, to the five years prior to its introduction. They found a long-term suppression of the pink bollworm (*Pectinophora*), one of the specific targets of Bt cotton, in Arizona. Comparing the rate of capture of adult moths in areas with high use of Bt cotton to those of low use within the state it is clear that the population is strongly trending downward in areas of high Bt use. The bollworm is a specialist feeder that lacks many alternate hosts, unlike some other insects that are cotton pests, so that it may be a somewhat special case. In this

instance, the decreased use of insecticides benefits overall biodiversity by reducing loss of non-target insects

A more recent study by Cattaneo, Yafuso, Schmidt, Huang, Rahman, Olson, Ellers-Kirk, Orr, Marsh, Antilla, Dutilleu, and Carriere, (2006) considered the impact of Bt cotton on biodiversity, at a farm scale, evaluating 81 fields across Arizona. Different fields were examined in each of two years, 2002 and 2003. Yields for the Bt cotton were similar to those of non-Bt cotton with applied insecticide. The non-Bt cotton received about six applications in both years, but broad-spectrum insecticide was applied more often in 2003 (~ 4 times) than in 2002 (~ 2 times) for the Bt cotton. When samples were collected in cotton fields compared to adjacent uncultivated areas, decreases were observed for ants species richness in cultivated fields, whether or not Bt cotton was the type being grown. For beetles, cultivated fields had greater densities and species richness. A complicating factor was the significant difference in soil types in areas with Bt cotton versus those without, which could have affected ant species diversity. Some of the Bt cotton also carried herbicide resistance genes. This additional trait showed no effect on the measured richness or diversity for ants or beetles. The authors concluded that Bt per se had no effect on species diversity.

Some of the most thorough, long-term studies of Bt effects were reported by Naranjo in two publications (2005a, b). The studies were carefully designed to provide high statistical power by using moderate to large plots and four replications, over several years. Comparing unsprayed Bt to non-Bt cotton, there was generally a decrease in 0-2 of 22 taxa of natural enemies on the Bt cotton, in any one year. Over the whole study there was about 19 % decrease in five predator taxa on the Bt cotton, presumably because of a dearth of prey species. Insecticide use resulted in much larger decreases in abundance of more taxa, about 48 % in 13 taxa. By having multiple years and replicates it was possible to have 80 % confidence in as little as a 20 % change in density. No long-term *trend* could be seen in predator populations comparing Bt to non-Bt cotton, although there was a slight decrease of population density (~ 6 %) in side by side comparisons.

In addition to abundance (Naranjo 2005a) the function of the community is of importance (Naranjo 2005b). Bt had no effect, whereas insecticide use was detrimental to predation by various predators on tobacco whitefly (*Bemisia tabaci*), the pink bollworm (*Pectinophora gossypiella*) or lygus bug (*Lygus hesperus*). Averaged over all five years, there was a decrease of predation index signifying a decrease in effective predator: prey ratios, to the extent of about 30 % for whitefly, and 50 % for the other two species. The predation indices were derived from previous studies of prey preference for various predators and are thus robust measures.

Mortality of bollworm eggs supplied as "egg cards" was tested in the field as a direct measure of predation activity. Pupae were also tested in the field. For eggs, predation was about 40 % in 24 h while for the pupae levels were nearer 60 %. There was no effect of Bt vs non-Bt (unsprayed) over several experiments in two or three years. Predation or parasitism on various life stages of the

whitefly showed no effect of Bt vs non-Bt (unsprayed) cotton. For the lygus bug no comparable direct experiments could be done.

Although there were decreases in some predator populations on Bt cotton (Naranjo 2005a) it was concluded that compensatory activity by other predators made up for the small difference seen. Broad spectrum insecticide treatment on the other hand is well known to lead to outbreaks of some pests such as aphids through damage to the predator population. Sisterton, Biggs, Manhardt, Carriere, Dennehy and Tabashnik (2008) document the benefit that reduced insecticide use has on populations of generalist predators in Bt cotton.

7.2.3 Irrigated Australian Cotton

Whitehouse, Wilson and Fitt (2005) made a detailed comparison of the whole insect community in canopies of Bt and non-Bt cotton on commercial Australian cotton farms. Plot sizes were all above 1 ha, with replications for the unsprayed treatments. In the first season (1995/1996) there were two locations. One of those locations was used a second season and a different location was used the third year. From 43 to 88 taxa were identified in the four experiments. There were wide ranges in numbers of specimens recovered on different sampling dates, and small numbers of many species per sample so that all samples per date (9 or 10) were combined for some analyses. Community structures changed over time, with date of sampling accounting for about half of all the variance and crop type accounting for 9-16 %. The sprayed conventional cotton was the most different from the others, with lower populations on most dates, as expected. The Bt cotton lowered numbers of bollworms as anticipated but also was associated with changes in numbers of several other taxa, particularly predators that depend on the bollworm as a food source. Unsprayed conventional cotton had a slightly greater diversity than unsprayed Bt cotton but different diversity indices gave different results, at least to the level of significance. Overall the effect of crop type was small ($< 5\%$ of variance).

7.2.4 Relay-cropping in North China

Two papers describe the potential impact of transgenic cotton on insect pest and predator populations in northern China (Men, Ge, Liu and Yardim 2003; Men, Ge, Edwards and Yardim 2004). The experimental design approximates the actual on-farm situation, with non-isogenic cotton lines being compared. Plot sizes were large, at 0.4 ha (~ 1 acre), and laid out in a randomized complete block design with four treatments and three replicates, studied over three years. The large plot size helps avoid spray drift between treatments. The transgenic (Bt) line was Deltapine NuCOTN 33B, while the local variety of similar appearance and growth habit was Chun Aizao. These are not isogenic, or near-isogenic lines. Seedlings were transplanted into a field during May with the winter wheat crop harvested in June. Predators or pests may over-winter in the alternate crop. Timing of insecticide applications was based on the appearance of

threshold levels of pests, in an integrated pest management strategy. For the Cv Chun Aizao, insecticide use included treatment for bollworms in two seasons. The Cv NuCOTN 33B required treatments for mites, mirids and leafhoppers in seasons when Chun Aizao did not. Both lines received aphid treatment on one or two dates each season. Predators including ladybeetles and spiders were the focus for Men et al (2004), while overall diversity was the focus of Men et al (2003). Sampling was by visual inspection of 6 plants ($\sim 1 \text{ m}^2$) in 5 locations per plot for sedentary arthropods and 5 sweeps of a net for flying ones. Aphids were counted directly on three leaves per plant.

As expected, insecticide use had a strong negative impact on spiders, but little effect on numbers of ladybeetles, which may have migrated in from surrounding untreated areas (Men et al 2004). Insecticide application paradoxically increased the Shannon-Weaver diversity index on Bt cotton, by increasing evenness through reduction in populations of more abundant arthropods (Men et al 2003). Overall, three dozen species (16 predators and 20 pests) were observed, but many of them were sparsely represented. Fourteen species were found in only 1/4 of sampling dates and the average number of species for most treatments and years was below 20. Overall, eight insecticide applications were used on Chun Aizao in three years, while NuCOTN 33B required nine. Thus labor and energy costs were not reduced by the introduction of Bt cotton. However it appears that the IPM criteria for insecticide application are much lower than the economic thresholds commonly used in the U.S., for instance in California.

Men et al (2004) did not comment on the striking result that aphid populations crashed on untreated plants at nearly the same time as those on insecticide treated plants, for both Bt and non-Bt lines. Such a population crash suggests that there was some natural control occurring. This could have been from predation, or more likely, fungal infection of the aphids (Feng, Chen and Chen 2004). A number of fungi infect aphids in a population dependent manner, often producing a population crash. In the U.S., monitoring of level of fungal infection is a part of a successful integrated pest management strategy in the southeastern cotton growing region. There it is recognized that early application of insecticides results in explosions of the aphid population (Long et al 2002). A weakness of the studies of Men et al (2003, 2004) compared to studies cited above, is the relatively low intensity of sampling.

A less dramatic shift in insect pests was noted by Smith (1998) during the final stages of boll weevil eradication in Alabama. The introduction of Bt cotton reduced insecticide applications from 6-12 down to 0-4, which in turn allowed build-up of plant bugs (*Lygus*), armyworms and stink bugs later in the season. In earlier years those had not been major concerns compared to boll weevils and bollworms. Such secondary pest outbreaks in China have resulted in increased pesticide applications, negating some of the savings that come from use of Bt cotton (Just, Wand and Pinstrup-Anderson 2006). This is discussed more extensively below.

7.2.5 CpTI (cowpea trypsin inhibitor) in China

China is the only country where an insect resistance in cotton other than Bt is deployed (Li, Wu, Chen, Feng, Xu, and Guo 2004). Li et al studied one Bt line that also carried a cowpea trypsin inhibitor. The experimental design used comparison to parental Cvs, unlike the work of Men et al (2004), so that effects ought to be more closely related to the effects of single genes, rather than the whole genotype. Plots were small at 3.3 x 3.3 m with about 50 plants in each (45,000/ha). Each year there were three treatments; the parental line, parental line sprayed, and Bt line, with three replications. The only insecticide used was a pyrethroid, applied up to 15 times per season to control bollworm. A suction sampler was used, apparently on 10 plants per plot every 10 days. The Shannon-Weaver diversity index was used as a measure of biodiversity. There were no significant differences between treatments, when pest, natural enemies, or neutral arthropods were considered, and only a couple of dates and treatments where there was an effect on the community as a whole. Unfortunately absolute numbers are not indicated anywhere. It is likely they were rather small so that the statistical test of the Shannon-Weaver index had little power.

7.2.6 The Greenpeace Report in China

Adoption of Bt cotton was rapid and extensive in China. By 2002 Greenpeace had published a report prepared by D. Xue, emphasizing potential negative impacts of Bt cotton adoption. The report received considerable world-wide attention from the press. The actual report is a thorough and reasonably fair summary of findings, from many sources, including unpublished reports. Only a few highlights are given here; original citations are provided in Xue (2002).

Natural predator populations were much higher in Bt cotton than in conventional cotton with insecticides, but lower than in untreated conventional cotton in a 1998 study of Wu, in Henan Province, northern China. On the other hand Cui and Xia at the same time in a different area of the same province found no difference in predator numbers, but noted that bollworm parasitoid populations were decreased on Bt cotton. Shen in 2001 also found decreased parasitization rates on bollworm larvae in Bt cotton fields in Jiangsu Province, eastern China. Zhang working in Hubei (humid, rainfed, central China), and Wei working in Xinjiang (irrigated, northwest China) both found no significant difference in predator populations on Bt vs conventional cotton.

Among secondary pests of cotton, beet armyworms were not fully controlled by Bt cotton, in a Henan study by Wu (1999). (In the U.S., there is no expectation of control of such moths by Bt). Lygus bugs were not controlled at all by Bt in a later Hebei study of Wu (2001) nor would control be expected. Numbers exceeded generally recognized economic thresholds for U.S. production (MS State 2007b) with up to 40/ 100 plants in 1999 and 320/ 100 plants in 2000. Whiteflies increased late in the season in the same study area. Studies by Cui

and Xia in Henan showed increases of some secondary pests on Bt cotton, including aphids, spider mites, whiteflies and thrips. Zhang examined crops in humid, southern Hubei in 1999 and found some increase of cotton aphids. An important factor likely to affect the outcome of both these studies is the same feature noted above in the work by Men et al (2003, 2004), namely the failure to study isogenic lines of Bt and non-Bt cotton. Zhang reportedly concluded that it was the Cv rather than treatment that affected population levels for aphids. The Monsanto lines widely adopted in China were not bred specifically for either their several climatic zones, or for some of the cropping systems used there. So they may well differ from conventional cotton in their response to and support of various pests.

The concept of economic threshold was not addressed in any of these studies and the total numbers of insects per plant were often quite low by U.S. standards. For instance in Zhang's Hubei work, aphids peaked at 18 per plant (3rd leaf from top) and red spider mites peaked at 27. In California, 50 aphids per 5th leaf is near the mid-season economic threshold but at the end of the season thresholds are lower, when sticky cotton can result from exudates. For spider mites it is the fraction of plants infested that is critical to control or economic impact. Similarly, for thrips in Cui and Xia's work the numbers attained were very low until late in the season when they would have little impact at the observed level of 3-4/leaf. Zhang, in Xinjiang did find thrips at ~10 fold higher levels than this. Whiteflies never reached 8/leaf in the study of Cui and Xia, and for most cases were below 4/leaf although they were present throughout most of the season.

Looking more broadly at the range of species present, Zhang found in Xinjiang during 1999 exactly the same total predators +pests (38 species), but only 8 species of spiders on Bt cotton vs 19 on non-Bt cotton. A study by Wei and Zhang in northern China (Hebei Province), supported by Monsanto, noted 41 species of arthropods in each of two kinds of Bt cotton, 42 species in insecticide-treated conventional cotton and 49 species in untreated conventional cotton. The one striking result of this study was that total arthropod numbers were about 4-fold higher in the treated conventional cotton, suggesting an outbreak of some kind of insect, identity unspecified.

7.2.7 Responses to the Greenpeace Report in China

A detailed rebuttal of the critical comments in the 2002 Xue (Greenpeace) report is found in a review by Jia and Jin (2004). Those authors cite a number of direct measures, principally obtained by the authors themselves or their colleague K.M. Wu, that have been published in a range of venues. Surveys indicated a 70-80 % reduction in pesticide use in Bt cotton, and a decrease of cotton aphids, which they attribute to increased predator numbers. They also claimed that arthropod biodiversity was increased (by an unspecified amount) in Bt cotton fields.

A directly relevant study was done in Hubei Province, in central China, which has a rather different climatic regime (much more humid) than the major cotton-producing areas of the northwest provinces. The study (Deng, Xu, Zhang, Zhou and Xu 2003) examined three treatment conditions in two seasons. Conventional cotton with IPM was used as reference. In Bt cotton fields, aphids and spider mites rose about two-fold compared to those of the IPM conventional cotton, but spiders and other predators (*Propylaea japonica*, a ladybeetle, and *Geocoris pallidipennis*, a bigeye bug) rose by similar amounts.

The issue of secondary pests in Bt cotton in China was emphasized again in a study, done in 2001 through 2004, and reported at an agricultural economics meeting by workers associated with Cornell University (Just, et al 2006). Their press release strongly suggested that secondary pests might reverse the economic gains of Bt cotton, based on finding that many farmers were applying large amounts of insecticides to control secondary pests, and paying high prices for Bt seed. However Hu, Huang, Lin, Rozelle and Pray (2006) showed that the increase in insecticide use in a sample of villages across the cotton growing area was only a small fraction (3.4%) of the reduction obtained by use of Bt cotton, and that the pressure from such secondary pests decreased from 2004 to 2005. Their conclusion was that there is no evidence that the problem of secondary pests is becoming progressively worse *over time*. A fuller version of their work is available as a working paper (Huang, Lin, Hu, Rozelle and Pray 2006). It should also be pointed out that institutional failures and sociocultural factors may be more responsible for the increased insecticide use (Stone 2004), than are economically significant outbreaks.

Many other studies of Bt cotton have been done in China. Wu and Guo (2005) provide a clear review of some that are not readily accessible. As described by those authors, Bt resistance had not risen in the primary target pest *Heliothis armigera*, because Bt corn was not yet (2005) in use and it provided a large refuge area for the insect (~ 10 -fold by area), even when only Bt cotton is planted, as in Shandong, Hebei and Henan provinces. Also, long distance migration of the moths may constantly reintroduce susceptible types into areas where reproduction is strongly suppressed on cotton by the presence of Bt genes. There is concern that if Bt corn is introduced, the pressure on pests may lead to a rise of resistance (Wu 2007).

7.2.8 Bt Introduction and Institutional Failures in India

The introduction of Bt cotton aroused great controversy in India (e.g. Grain 2007a), but it has proceeded very rapidly nevertheless. The estimated area planted to Bt cotton now exceeds 6.2 million ha, about 2/3 of the total cotton acreage, and there are now more than three dozen Cvs available (Geeta 2007, ISAAA 2008). The first Bt cotton grown in India was not legal, and there were very few lines available for the first couple of years after its legal introduction. These had a limited range of adaptability, and were hybrids with the Bt gene brought in from only one parental line rather than true-breeding lines.

A different life cycle than used in U.S. production (longer growing season with harvest of two flushes of bolls) led to difficulties, particularly in some regions, because the content of Cry1 Ac decreases in late season and the concentration is relatively lower in bolls than in leaves (Kranthi, Naidu, Dhawad, Tatwawadi, Mate, Patil, Bharose, Behere, Wadaskar and Kranthi 2005). Institutional failures were widespread (Jayaraman 2005 & 2005, Government of India 2006, Grain 2007a).

A "farmer's participatory field trial" was done in central India as described by Bambawale, Singh, Sharma, Bhosle, Lavekar, Dhandapani, Kanwar, Tanwar, Rathod, Patange and Pawar (2004). A single line of Bt cotton (Bt-MECH-162) was used under integrated pest management (IPM) by comparison to conventional cotton (some was MECH-162 which provided refuges for lepidopterans) with and without IPM. Seven sprayings of insecticide were used on the conventional cotton, yet it sustained over 50 % damage to bolls by bollworms, compared to ~30 % in IPM conventional or non-Bt-MECH-162 and only <12 % in Bt-MECH-162. Pink bollworm infestation showed a similar difference of 24 % in sprayed conventional vs < 4 % in IPM Bt-MECH-162. Aphids and thrips were significantly increased in the sprayed conventional cotton. Numbers of predators, measured as lacewing eggs and ladybeetle adults, were significantly lower on the MECH-162 Cv compared to conventional unsprayed, but higher than on sprayed. The Bt-MECH-162 and non-Bt-MECH-162 showed identical counts of beneficial insects, suggesting a Cv effect, rather than a dearth of prey. Because none of the experimental fields are true random replicates, one cannot exclude some environmental (site) effects.

Bambawale et al (2004) mention one important example of an institutional failure. They noted that in a nearby village, farmers were spraying their Bt cotton because they saw bollworm eggs on it, not recognizing that the Bt effect occurs only after the eggs become larvae. Another problem that they identified was susceptibility of MECH-162 to fungal wilt (see discussion below).

Additional problems are documented in a report by the Asia Pacific Consortium on Agricultural Biotechnology (APCoAB 2006). These included much use of undocumented Bt lines, saving of hybrid seed which leads to 1/4 of the offspring being fully susceptible, and inappropriate claims for the efficacy of Bt in control of insects other than lepidopterans. No evidence has been found of alterations in agroecosystem function caused by the Bt genes, but there are significant Cv background differences which do affect other pests and predators.

Some of the Bt lines are more susceptible to fungal pathogens including wilt than are conventional adapted Cvs (Government of India 2005). There are many anecdotal reports that cotton grown without pesticide in India shows better productivity than Bt cotton, because the available Bt cotton Cvs were susceptible to insects other than those controllable by Bt (Gala 2005). These disadvantages of Bt cotton represent institutional failures, in rushing to market lines that are not well adapted for the prevailing cultural conditions. There is no evidence that the Bt gene itself increases susceptibility to any disease or pest.

Sharma and Pampapathy (2006) documented differences in susceptibility of Bt cotton to some non-target pests. In particular the jassid bug (leafhoppers) was a problem on some of the transgenic lines, but also on their progenitor lines, as compared with other hybrids or *Gossypium arboreum* Cvs. This indicates one of the problems with a particular breeding strategy; if it is expected to have benefits that it cannot provide, it will fail.

As discussed by Stone (2004) biotechnology in which only single traits are transferred may be particularly problematic when rapidly introduced to a culture. Identity confusion, misinformation, and unrealistic expectations are all common problems when transferring portions of a technology to different social and cultural agricultural communities.

7.2.9 Status of Bt Cotton in Other Countries

Bt cotton was only recently approved for use in Brazil. It has been present in Argentina legally for a number of years and very likely illegally in Brazil for nearly as long (Grain 2007). In 2007, about 500,000 ha of Bt cotton was grown in Brazil, nearly half the total area (ISAAA 2008). Almost all of the Argentine crop is Bt. There are no published reports of impacts on agroecosystems in those countries. Bt cotton was introduced to Indonesia but crop failures led to withdrawal of the seed supplier (Down to Earth 2004).

In Mexico, about 1/3 of the cotton is Bt, because the bollworm is a major pest in only some regions. Implementation of Bt cotton use has been very successful in that country with significant economic benefits accruing to farmers (Traxler and Godoy-Avila 2004). No reports on agroecosystem effects of Bt cotton are available.

Uzbekistan, one of the larger cotton producers, and second largest exporter (USDA ERS 2008), has not yet adopted Bt cotton (Grain 2007). It includes the most northern cotton-growing latitudes and needs specially selected, short season Cvs for successful irrigated production, comparable in day length response to those used in Xinjiang Province of China.

Pakistan, unlike its neighbor India, was slow to adopt Bt cotton. Large quantities of illegal Bt seed are in use, perhaps 80 % of the crop, but there are still no officially approved Cvs. Those that are in use, reportedly carry, or are susceptible to cotton leaf curl virus (Daily Times 2008). As mentioned above, yields are relatively low and very large areas (3.2 million ha) are needed to produce the 9-11 million bales (2 million metric tons) estimated for 2007. With no acknowledgment of the use of Bt cotton, there are no reports of its direct or indirect effects.

Amongst countries on the African continent, South Africa has the most advanced Bt cotton program with over 80 % of production containing the Bt gene. Raney (2006) has summarized economic benefits and institutional failures. Introduction of Bt cotton in 1998 led to a reduction of pesticide use with benefits to both large and small scale production which coexist. Institutional

factors more recently led to a massive decline in small-scale rain-fed production. No negative agroecosystem effects have been documented for Bt cotton. One extensive study over two seasons comparing Bt and non-Bt cotton documented that spider populations were not negatively affected (Mellet, Schoeman and Dippenaar-Schoeman 2006). It was also observed that spraying with endosulfan did not diminish spider populations on non-Bt cotton. A further study over the same time period and location examined egg parasitism on bollworms (Mellet and Schoeman 2004). Until very late in the season, the extent of parasitism on Bt cotton was the same or better than for non-Bt with or without endosulfan spray which was used to control the pest when larval numbers reached an economic threshold of five/ 24 plants. On Bt cotton the bollworm never reached an economic threshold; in fact it never exceeding 1.6/ 24 plants. The authors suggest that so long as refuges are provided for the bollworm, the parasite will remain a useful controller of potential outbreaks.

8 Herbicide Resistance

Amongst herbicide resistance strategies, resistance to glyphosate, first produced by introduction of a bacterial gene essential for aromatic amino acid biosynthesis, is the most prominent. Indeed, glyphosate is one of the most widely used herbicides world-wide. Some additional tools have been developed to enhance glyphosate resistance in modified plants, including oxidation and acetylation of glyphosate. While the glyphosate oxidase gene is deployed in canola and maize, it is not listed in a database of glyphosate resistance mechanisms for cotton (Agbios.com 2007). Presumably there is only one mechanism of resistance being used in cotton.

Scientists surveyed by Culpepper (2006) indicated that a number of weeds, including amaranth, sedges, bindweed and tropical spiderwort, had become more troublesome as the extent of adoption of glyphosate-resistant cotton was increased. The most disconcerting example is the tropical spiderwort (*Commelina benghalensis*) which has high natural tolerance to glyphosate and which becomes more abundant in no-till systems. This plant has been identified as the worst weed in cotton in Georgia recently (Prostko, Culpepper, Webster, and Flanders 2005) although it was not a notable weed at all a decade earlier.

Resistance to a second herbicide, glufosinate, has been commercialized more recently under the trade names of Liberty or Basta. In this case a prokaryotic gene (PAT), from a streptomycete, was introduced into cotton to allow it to detoxify L-phosphinothricin (the active ingredient of glufosinate) by acetylation. Plants carrying the PAT gene are able to tolerate direct application of the herbicide while in advanced stages of growth. Dow (2007) has used the PAT gene as a selectable marker in development of Widestrike cotton which has two Bt genes. A copy of the PAT gene comes from each of the parental lines donating a Bt gene. Many annual weeds are effectively controlled by

glufosinate, but some grasses and the dicot amaranths are poorly controlled under many conditions. Webster, as cited by Gardner, York, Jordan and Monks (2006) indicated that the Palmer amaranth was the second worst weed of cotton in the southeastern U.S. by 2005. Earlier studies by Coetzer, Al-Khatib and Peterson (2002) indicate that amaranths are relatively resistant to glufosinate once they attain a size of a few inches (~10 cm) in height. Thus Gardner et al (2006) conducted studies of glufosinate-resistant cotton with various combinations of pre- and post-emergent herbicide applications. They looked at six locations and four herbicides applied pre-emergence, with glufosinate applied for post-emergence control. In one case both the glufosinate and another herbicide (pyrithiobac) were applied post-emergence. Amongst the important conclusions of this study, it is apparent that the PAT transgene and glufosinate alone are not sufficient for weed control, in North Carolina.

Use of glufosinate alone could result in a marked shift in weed populations toward more resistant types. Also, it is not clear that the glufosinate resistance of cotton allows for a significant reduction in *number* of seasonal herbicide applications needed. Glufosinate is used on only a small fraction of cotton at present, so “selected” resistance has not been reported as a problem with weed control. If Wide-strike cotton is adopted, glufosinate usage might increase considerably.

Bromoxynil-resistant cotton was the first transgenic cotton. It has not been a commercial success, and is no longer in use in the U.S., at least (Robinson 2004). Initially, concern was expressed that residual herbicide might contaminate the seed and gin trash which could end up in cattle feed and make it's way through the food chain to people via milk or meat (USEPA 1998). Thus the permitted acreage was limited to a small fraction of the total planted to cotton. Its ultimate downfall was that it was not available in many genetic backgrounds, meaning that it could not be widely adopted, and it did not provide for effective control of as many weeds as glyphosate (Robinson 2004).

9 Disease Resistance

Two wilt diseases, induced by *Verticillium* and *Fusarium* species, may have major impacts on cotton production particularly under stress conditions. Root knot nematodes somehow interact with the *Fusarium* wilt disease, so that nematode-infected plants are more susceptible to the wilt (Davis, Colyer, Rothrock, and Kochman 2006). Resistance to the *Verticillium* wilt has been successfully developed by conventional plant breeding, but less progress has been made with *Fusarium* for which there are many races, which have alternate hosts, and which can survive as spores in the soil for extended times.

An anti-fungal protein from a traditional Chinese medicinal herb, when over-expressed in colored cotton lines, increases resistance to the *Verticillium* wilt (Wang, Chen, Wang, Huang, Yao, Liu, Wei, Li, Zhang, and Sun 2004).

Three different Cvs were treated using the “pollen-tube pathway” of seed transformation with plasmid DNA carrying the desired gene, plus a herbicide-selectable marker (BAR gene). Two effective, stable transformants were obtained from over 18,000 seedlings derived from over 16,000 flowers. The efficiency of the pathway is thus very low. Nevertheless, this is a first report of transgenic fungal resistance in cotton. The resistant lines were field tested into the F₂ generation. There is no obvious ecosystem impact of such a modification of the line, per se, because it is equivalent in resistance to commercially available, conventionally bred lines, although those conventional lines are not adapted to the growing conditions of interest to the authors (northwestern China). There is no evidence that the BAR gene or the *Verticillium* resistance have impacts beyond the productivity of the cotton crop. The BAR gene is equivalent to the PAT gene and can confer resistance to glufosinate, but that herbicide resistance was not tested in the field by Wang et al (2004).

There have been other transgenic, fungal resistant cotton lines produced but not yet refined to the level of field testing. Chitinase genes, from beans (Tohidfar, Mohammadi and Ghareyazie 2005) or *Tricoderma virens* (Emani, Garcia, Lopata-Finch, Pozo, Uribe, Kim, Sunilkumar, Cook, Kenerley and Rathore 2003) have been transformed into cotton in culture with plantlets selected and regenerated into full size plants. Gene expression and fungal resistance have been demonstrated through at least the T₁ generation.

10 Potential for Gene Flow from Transgenic Cotton

The above sections considered how transgenic cotton might alter organisms that interact with the cotton plant. Here, the potential for altering the gene pool of wild or domesticated species is considered. In many countries, agronomic cotton is grown far from native relatives, or interfertile native species are not present, so that concerns for gene flow deal not with potential for affecting wild relatives, but rather with the issue of maintaining genetic identity of coexisting cotton crops (Hutmacher, Vargas and Wright 2006). As for major food crops, some importing countries may have restrictions on transgenic cotton products. For instance, countries in the European Union have strict limits on the level of detectable presence of transgenic types in soybeans, wheat and corn (usually 0.5 % of the seed or material). When a transgene's presence is not desired, it is called contamination. Similarly, cottonseed meal produced from transgenic plants would be contaminated with insecticide or herbicide resistance genes, and probably also by their respective proteins. Hence if one wishes to produce certified organic cottonseed or cotton fiber, it is important to maintain the purity of the lines. Pollen is the most likely source of unwanted contamination for stock seed, although there is potential for contamination during seed handling by the seed supplier.

An exemplary study of pollen-mediated gene flow was described for California (Van Deynze, Sundstrom and Bradford 2005). Herbicide resistance genes were monitored by bioassay of seedlings, and confirmed by DNA analysis. During a two year study, gene flow was similar in all directions, and declined exponentially away from the experimental plot, to less than 8 % at 1 ft (0.3 m) and < 1 % at 30 ft (9 m), when there were honeybees active. The rate of decline was 9-fold more rapid without the pollinators. With pollinators present, sporadic gene flow was detected up to 1 mi (1.6 km) but only to a level of 4/10,000. This is well below the EU criterion for contamination.

Similar studies of other areas are cited by Hutmacher, et al (2006) in a University of California publication which reviews agricultural biotechnology. Hutmacher et al (2006) directly address the distances needed to maintain seed purity, both for Foundation seed production and for maintenance of organic certifications. Using observation of fields by California Crop Improvement Association inspectors, they noted that separation distances of ~1.6 mi (~2.6 km) still showed occasional hybrids at a frequency near 1/10,000 which is the limit for Foundation seed. Similar separation distances may be needed for certified organic cotton.

The potential for modern cotton to transfer genes to wild relatives in Australia has been thoroughly discussed (Gene Technology Regulator 2002). Because cultivated cotton is an allotetraploid, originally from Mexico (Brubaker and Wendel 1994), while the native Australian species are diploids, the success of deliberate or spontaneous pollinations is very low with those species. Occasional triploids can be produced, but functionally this route is a dead end for gene transfer (Brubaker and Wendel 1994; Brown, Stewart, Kilby and Grace 1999). Surprisingly, there seem to be no articles considering transgene flow from domesticated cotton in Mexico, although Mexico was an early adopter of Bt cotton, and wild progenitors and primitive land-races are present (See Chap 1, this volume)

11 Sociopolitical Considerations and the Advance of Genomics

In several sections above, reference was made to institutional failures, or controversies regarding introduction of transgenic cotton. A fuller discussion is provided for the whole range of biotech crops in a recent review (Davis 2006). As discussed by Stone (2004) there are significant social constraints on introduction of an advanced technology with complex stewardship requirements in a culture with a diverse, fragmented agricultural system such as is found in most African countries. A specific instance for Bt cotton is discussed by Hofs, Fok, and Vaissayre (2006). In that setting of smallholders in South Africa, Bt cotton did not reduce overall pesticide use very much, labor savings were relatively small, and technology costs further reduced the marginal benefit to the farmers.

One general concern with crops engineered to show resistance to insects is the strong evolutionary pressure for insects to develop resistance to the toxin(s) of the plant. Tabashnik et al (2008) document the effectiveness of refuge strategies in delaying an increase in resistance alleles within the insect population. Global monitoring has shown increased frequencies of resistance alleles in *Helicoverpa zea*, but thus far a crop failure has not resulted, because alternative strategies including use of multiple toxin genes has been successful, and provision of refuges has delayed appearance of resistance at high frequency.

Economic costs associated with regulatory approval have proven a strong constraint on introduction and success of biotech crops, even in countries where the agro-ecosystem is intensively managed as in the U.S. Many potential applications of transgenic cotton are discussed elsewhere in this volume. Whether these can move beyond the stage of preliminary field testing and into wide scale commercial application remains to be determined. At this stage of world social development, their successful deployment will only happen if they provide a large increment in the final harvested crop, and a means by which the developers of the technology can recoup their investment with a significant profit margin. The many articles cited by Geeta (2007) for India, the Xue- Greenpeace report (Xue 2002) for China, and Raney's (2006) analysis for developing countries, can be consulted for detailed examples of the institutional challenges presented to deployment of biotech crops around the world.

Genomic approaches with plants have rapidly provided us new insights into plant growth and development. Now that genomes of several dicot species are sequenced to a several-fold level of redundancy, patterns of gene expression, chromosomal organization and evolution are becoming more apparent. Comparative genomics is still in its infancy, but it is already clear that we must obtain sequence information for the crop of interest in order to understand the key regulatory differences that make it a crop and not simply another weed. Saxena and Brown (2005) indicated that for the production of cellulose fibers, far more than just the Cesa gene family would be involved. Transcriptome analysis (e.g. Chapter 7) indicates that hundreds of genes show differential regulation in the cotton fiber cell.

Some of the studies cited above show that there are significant Cv differences in important environmental responses such as insect resistance. Some of these are conditioned by altered levels of secondary products such as tannins and gossypol. Genomic studies may help give a molecular basis to such alterations, which depend on altered fluxes through biosynthetic pathways. This, in turn, may allow engineering of improvements.

However, the institutional challenges to deployment of biotech crops are large and are likely to remain. It thus appears likely that only a few of many possible incremental improvements will prove economically viable. The best way forward may be stacking several useful traits, each of which provides an incremental benefit, and clustering them together on a single chromosomal region, e.g. via a T-DNA insertion, so that the entire stack of traits may be simultaneously transferred by conventional breeding into well adapted Cvs.

For instance, several gene alterations that enhance cotton fiber length and strength might be combined with those for increased metabolic flux to the site of secondary cell wall synthesis (Haigler et al 2007), in a way that allows their enhanced expression in the trichomes of the coat of the developing seed. Issues of regulatory approval of transgene events, versus final traits, would need to be resolved for such a strategy to work. At this point we are in the realm of speculation, but at the rate that advances are occurring in model systems, it may be feasible within a decade.

Acknowledgments This is contribution 07-304-B of the Kansas Agricultural Experiment Station. I thank Andrew H. Paterson for advice and encouragement.

References

- Agbios.com (2007) available at <http://www.agbios.com/>
- Ahmad, M. (2007) India outpaces Pakistan in boosting cotton yield. Southeast Asia Biosafety Program News, Thursday March 27, 2007 available at http://www.agbios.com/sabp_-main.php
- APCoAB (2006) Bt cotton in India. A status report. Asia-Pacific Consortium on Agricultural Biotechnology, ICRISAT, New Delhi, India, 34 pp, available at http://www.parc.gov.pk/bt_cotton.pdf
- Bambawale, O.M., Singh, A., Sharma, O.P., Bhosle, B.B., Lavekar, R.C., Dhandapani, A., Kanwar, V., Tanwar, R.K., Rathod, K.S., Patange, N.R., and Pawar, V.M. (2004) Performance of Bt cotton (MECH-162) under integrated pest management in farmer's participatory field trial in Nanded district, central India. *Curr. Sci.* 86, 1628–1633.
- BASF (2007) Clearfield technologies, available at <http://www.agsolutions.ca/basf/agprocan/agsolutions/WebASCclearfield.nsf/defaultWest.htm>
- Brubaker, C.L., and Wendel, J.F. (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Amer. J. Bot.* 81, 1309–1316
- Brubaker, C.L., Brown, A.H.D., Stewart, J.M., Kilby, M.J., and Grace, J.P. (1999) Production of fertile hybrid germplasm with diploid Australian *Gossypium* species for cotton improvement. *Euphytica* 108, 199–213
- Burcher, S. (2007) Picking cotton carefully, Press release Mar 15, 2007 of Institute for Science in Society, London, UK available at http://www.i-sis.org.uk/Picking_Cotton_Carefully.php
- Canada (2006) Decision documents- determination of environmental and livestock feed safety. Canadian Food Inspection Agency- Plant Biosafety Office, available at <http://www.inspection.gc.ca/english/plaveg/bio/dde.shtml>
- Cantrell, R.G. (2006) The role of biotechnology in improving the sustainability of cotton, available at <http://www.cottoninc.com/Sustainability/>
- Carriere, Y., Ellers-Kirk, C., Sisterton, M., Antilla, L., Whitlow, M., Dennehy, T.J., and Tabashnik, B.E. (2003) Long-term regional suppression of pink bollworm by *Bacillus thuringiensis* cotton. *Proc. Nat. Acad. Sci. USA* 100, 1519–1523
- Cattaneo, M.G., Yafuso, C., Schmidt, C., Huang, C.Y., Rahman, M., Olson, C., Ellers-Kirk, C., Orr, B.J., Marsh, S.E., Antilla, L., Dutilleul, P. and Carriere, Y. (2006) Farm-scale evaluation of the impacts of transgenic cotton on biodiversity, pesticide use, and yield. *Proc. Nat. Acad. Sci. USA* 103, 7571–7576.
- Coetzer, E., K. Al-Khatib and D.E. Peterson (2002) Glufosinate efficacy on *Amaranthus* species in glufosinate resistant soybean. *Weed Technol.* 16, 326–331

- Cotton Council International (2007) Directories and Buyers Guide, Cotton Council International, available at <http://www.cottonusa.org/directories/>
- Culpepper, A.S. (2006) Glyphosate-induced weed shifts. *Weed Technol.* 20, 277–281.
- Daily Times (2008) MINFAL seeks approval from ECC on Bt cotton. Available at <http://www.dailytimes.com.pk>, search article title.
- Davis, L.C. (2006) Genetic engineering, ecosystem change and agriculture: an update. *Biotechnol. Mol. Biol. Rev.* 1, 87–102. available at <http://www.academicjournals.org/BMBR>
- Davis, R. M., Colyer, P.D., Rothrock, C.S., and Kockman, J.K. (2006) *Fusarium* wilt of cotton: population diversity and implications for management. *Plant Dis.* 90, 692–703
- Deng, S.D., Xu, J., Zhang, Q.W., Zhou, S.W., and Xu, G.J. (2003) Effect of transgenic Bt cotton on population dynamics of the non-target pests and natural enemies of pests. *Acta Entomol. Sinica* 46, 1–5.
- Donegan, K., Palm, C., Fieland, V., Porteous, L., Ganio, L., Schaller, D., Bucac, L., and Seidler, R (1995) Changes in levels, species and DNA fingerprints of soil-microorganisms associated with cotton expressing the *Bacillus thuringiensis* var kurstaki endotoxin. *Appl. Soil Ecol.* 2, 111–124.
- Dow (2007) Widestrike insect protection, Dow Agrosiences, available at <http://www.dow.com/productsafety/finder/ws.htm>
- Down to Earth (2004) Indonesia to sign biosafety instrument, Down to Earth 62, available at <http://dte.gn.apc.org/62AGR.HTM>
- Emami, C., Garcia, J.M., Lopata-Finch, E., Pozo, M.J., Uribe, P., Kim, D.J., Sunilkumar, G., Cook, D.R., Kenerley, C.M., and Rathore, K.S. (2003) Enhanced fungal resistance in transgenic cotton expressing an endochitinase gene from *Trichoderma virens*. *Plant Biotechnol. J.* 1, 321–336.
- Feng, M.G., Chen, C., and Chen, B. (2004) Wide dispersal of aphid-pathogenic Entomophthorales among aphids relies on migratory alates. *Environ. Microbiol.* 6, 510–516.
- Fitt, G., Mares, C., and Llewellyn, D. (1994), Field evaluation and potential ecological impact of transgenic cotton (*Gossypium hirsutum*) in Australia. *Biocontrol Sci. and Technol.* 4, 535–538.
- Frisbie, R.E., El-Zik, K.M., and Wilson, L.T. (Eds.) (1989) *Integrated Pest Management Systems and Cotton Production*, John Wiley, New York pp 157, 367, 409.
- Gala, R. (2005) Organic cotton beats Bt cotton in India, ISIS Press Release 5/5/05, Institute of Science in Society available at <http://www.i-sis.org.uk/OCBBCI.php>
- Gardner, A.P., York, A.C., Jordan, D.L. and Monks, D.W. (2006) Management of annual grasses and *Amaranthus* spp. in glufosinate-resistant cotton. *J. Cotton Sci.* 10, 328–338.
- Geeta, R. (2007) available at http://life.bio.sunysb.edu/ee/geeta/BTCotton_India.html
- Gene Technology Regulator (2002) The biology and ecology of cotton (*Gossypium hirsutum*) in Australia, Office of the Gene Technology Regulator, Government of Australia 30 pp available at <http://www.ogtr.gov.au/pdf/ir/biologycotton.pdf>
- Government of India (2006) Monitoring of the Bt cotton fields in Gujarat, Report of State Agriculture Department, Ministry of Environment and Forests, available at <http://envfor.nic.in/divisions/csuv/btcotton/srgj.pdf>
- Grain (2007a) Bt cotton, the facts behind the hype. GRAIN, Barcelona, Spain, available at <http://www.grain.org/seedling/?id=457>
- Grain (2007b) Field trials and commercial releases of Bt cotton around the world. GRAIN, Barcelona, Spain, available at <http://www.grain.org/research/btcotton.cfm?id=306>
- Hagerty, A.M., Kilpatrick, A.L., Turnipseed, S.G., Sullivan, M.J., and Bridges, W.C. (2005). Predaceous arthropods and lepidopteran pests on conventional, Bollgard and Bollgard II cotton under untreated and disrupted conditions. *Environ. Entomol.* 34, 105–114.
- Haigler, C.H., Singh, B., Zhang, D., Hwang, S., Wu, C., Caim W.X., Kang, W., Kiedaisch, B., Strauss, R.E., Hequet, E.F., Wyatt, B.G., Jividen, G.M., and Holaday, A.S. (2007)

- Transgenic cotton over-producing spinach sucrose phosphate synthase showed enhanced leaf sucrose synthesis and improved fiber quality under controlled environmental conditions. *Plant. Mol. Biol.* 63, 815–832
- Head, G., Moar, W., Eubanks, M., Freeman, B., Ruberson, J., Hagerty, A., and Turnipseed, S. (2005) A multiyear, large-scale comparison of arthropod populations on commercially managed Bt and non-Bt cotton. *Environ. Entomol.* 34, 1257–1266.
- Hequet, E.F. and Abidi, N. (Eds.) (2006) *Sticky Cotton. Measurement and Fiber Processing*. Texas Tech University Press, Lubbock, TX.
- Hofs, J.L., Fok, M., and Vaissayre, M. (2006) Impact of Bt cotton adoption on pesticide use by smallholders: A 2-year survey in Makhatini Flats (South Africa). *Crop Protect.* 25, 984–988.
- Hu, R., Huang, J., Lin, H., Rozelle, S. and Pray, C. (2006) Bt cotton in China: Are secondary insect infestations offsetting the benefits in farmer fields? (Abstract) International Consortium on Agricultural Biotechnology Research 2006 Conference, available at <http://www.economia.uniroma2.it/conferenze/icabr2006/abstract/file/Pray%20Rozelle.doc>
- Huang, J., Lin, H., Hu, R., Rozelle, S. and Pray, C. (2006) Eight years of Bt cotton in farmer fields in China: Is the reduction of insecticide use sustainable? Freeman Spogli Institute of International Studies, Stanford University, working paper, available at <http://fse.stanford.edu/publications>, title *Bt_Cotton_Insecticide_Use_September_2006.pdf*
- Hutmacher, R.B., Vargas, R.N. and Wright, S.D. (2006) Methods to enable the coexistence of diverse cotton production systems. *Agricultural Biotechnology in California Series*, Publication 8191, University of California Division of Agriculture and Natural Resources, available at <http://anrcatalog.ucdavis.edu/pdf/8191.pdf>
- ISAAA (2008) Global status of commercialized biotech/GM crops, Brief 37, 2008, available at <http://www.isaaa.org/resources/publications/briefs/37/executivesummary/default.html>
- Jayaraman, K.S. (2004) Illegal seeds overtake India's cotton fields. *Nature Biotechnol.* 22, 1333–1334.
- Jayaraman, K.S. (2005) Monsanto's Bollgard potentially compromised in India. *Nature Biotechnol.* 23, 1326.
- Jenkins, J.N. (1999) Host plant resistance in cotton and its value, In Wiseman, B.R. and Webster, J.A. Eds., *Economic, Environmental and Social Benefits of Resistance in Field Crops*, Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD pp 45–57.
- Jia, S.R., and Jin, W.J. (2004) The international debate on the biosafety of genetically modified crops: scientific review of several cases of debate. *Chinese J. of Agric. Biotechnol.* 1, 3–8.
- Johnson, J., MacDonald, S., McKeivier, T., Meyer, L., and Skelly, C. (2006) The United States and world cotton outlook, *Agricultural Outlook Forum*, Feb. 17, 2006 available by searching the title at <http://www.usda.gov/wps/portal/usdahome>
- Just, D.R., Wang, S., and Pinstrip-Andersen, P. (2006) Tarnishing silver bullets: Bt technology adoption, bounded rationality and outbreaks of secondary pest infestation in China., presented at the American Agricultural Economics Association national meeting July 25, 2006, available at <http://agecon.lib.umn.edu/>. search by title
- Kranthi, K.R., Naidu, S., Dhawad, C.S., Tatwawadi, A., Mate, K., Patil, E., Bharose, A.A., Behere, G.T., Wadaskar, R.M., and Kranthi, S. (2005) Temporal and intraplant variability of CryI Ac expression in Bt cotton and its influence on the survival of the cotton bollworm, *Helicoverpa armigera* (Huebner) (Noctuidae: Lepidoptera). *Curr. Sci.* 89, 291–298.
- Lee, M.K., Walters, F.S., Hart, H., Palekar, N., and Chen, J.-S. (2003) The mode of action of the *Bacillus thuringiensis* vegetative protein Vip3A differs from that of CryI Ab delta endotoxin. *Appl. Environ. Microbiol.* 69, 4648–4657.
- Li, W.D., Wu, K.M., Chen, X.X., Feng, H.Q., Xu, G. and Guo, X.Y. (2004) Effects of transgenic cotton carrying CryI A + CpTI and CryI Ac genes on diversity of arthropod communities in cotton fields in north China. *Chinese J. Agric. Biotech.* 1, 17–21.

- Long, J.L., Flint, S.G., Layton, M.B., and Steinkraus, D. (2002) Influence of boll weevil eradication on cotton aphid populations in Mississippi cotton: year four. Proceedings 2002 Beltwide Cotton Conference, available at <http://msucare.com/insects/cotton/01aphid.htm>
- Marvier, M., McCreedy, C., Regetz, J., and Kareiva, P. (2007) Meta-analysis of effects of Bt cotton and maize on nontarget invertebrates. *Science* 316, 1475–1477.
- Mellet, M.A., and Schoeman, A.S. (2004) Impact of Bt cotton on bollworm populations and egg parasitism. Information Systems for Biotechnology, Virginia Polytechnic Institute and State University, Blacksburg, VA, available at <http://www.isb.vt.edu/news/news04.Dec.html>
- Mellet, M.A., Schoeman, A.S., and Dippenaar-Schoeman, A.S. (2006) The effect of Bt cotton cultivation on spider populations in Marble Hall South Africa. *Afric. Pl. Protect.* 12, 40–50.
- Men, X., Ge, F., Liu, W., and Yardim, E.N. (2003) Diversity of arthropod communities in transgenic Bt cotton and nontransgenic cotton agroecosystems. *Environ. Entomol.* 32, 270–275.
- Men, X., Ge, F., Edwards, C.A., and Yardim, E.N. (2004) Influence of pesticide applications on pest and predatory arthropods associated with transgenic Bt cotton and nontransgenic cotton plants. *Phytoparasitica* 32, 246–254.
- MS State U (2007a) Cotton aphids: *Aphis gossypii*. Mississippi State University, Agricultural Extension Service, available at <http://msucare.com/insects/cotton/profile/cottonaphids.pdf>
- MS State U (2007b) Tarnished plant bug: *Lygus*. Mississippi State University, Agricultural Extension Service, available at <http://msucare.com/insects/cotton/tpb.html>
- Naranjo, S.E. (2005a) Long-term assessment of the effects of transgenic Bt cotton on the abundance of nontarget arthropod natural enemies. *Environ. Entomol.* 34, 1193–1210.
- Naranjo, S.E. (2005b) Long-term assessment of the effects of transgenic Bt cotton on the function of the natural enemy community. *Environ. Entomol.* 34, 1211–1223.
- Naranjo, S.E., and Ellsworth, P.C. (2003) Arthropod communities and transgenic cotton in the western United States: implications for biological control. First Intl. Symp. on Biol. Control of Arthropods, R. van Driesche, Ed. USDA Forest Service, Morgantown WV Publ. FHTET-03-05 pp 284–291
- NCSU (1999) Cotton: Friendly fungus helps control aphids. North Carolina State University Agricultural Extension Service, available at http://cipm.ncsu.edu/ent/Southern_Region/99Programs/arkansas.html
- Prostko, E.P., Culpepper, A.S., Webster, T.M., and Flanders, T. (2005) Tropical Spiderwort identification and control in Georgia field crops. Georgia Agricultural Extension Service Circular 884 available at <http://pubs.caes.uga.edu/caespubs/pubcd/C884.htm>
- Raney, T. (2006) Economic impact of transgenic crops in developing countries. *Curr. Opinion Biotechnol.* 17, 174–178.
- Robinson, E. (2004) Agribusiness: BXN cotton varieties to be retired. Delta Farm Press, June 2, 2004. available at <http://deltafarmpress.com/news/060204bxn-cotton-retiring/>
- Rui, Y.K., Yi, G.X., Zhao, J., Wang, B.M., Li, Z.H., Zhai, Z.X., He, Z.P., and Li, Q.X. (2005) Changes of Bt toxin in the rhizosphere of transgenic Bt cotton and its influence on soil functional bacteria. *World J. Microbiol. Biotech.* 21, 1279–1284.
- Saxena, I., and Brown, R.M. Jr. (2005) Cellulose biosynthesis: current views and evolving concepts. *Ann. Bot.* 96, 9–21.
- Sharma, H.C., and Pampapathya, G. (2006) Influence of transgenic cotton on the relative abundance and damage by target and non-target insect pests under different protection regimes in India. *Crop Protect.* 25, 800–813.
- Shetty, P.K. (2004) Socio-ecological implications of pesticide use in India, Economics and Political Weekly Dec 4, 2004 available at <http://www.epw.org.in/articles/2004/12/7985.pdf>

- Shoemaker, R., McGranahan, D., and McBride, W. (2006) Agricultural rural communities are resilient to high energy costs. *Amber Waves*, April 2006, U.S. Department of Agriculture, Economic Research Service, available at <http://www.ers.usda.gov/AmberWaves/April06/Features/Energy.htm>
- Sisterton, M.S., Biggs, R.W., Manhardt, N.M., Carriere, Y., Dennehy, T.J., and Tabashnik, B.E. (2007) Effects of transgenic Bt cotton on insecticide use and abundance of two generalist predators. *Entomol. Exp. Appl.* 124, 305–311
- Smith, R.H. (1998) Pest patterns. The impact of Bollgard technology and boll weevil eradication on cotton IPM in Alabama. *Highlights of Agricultural Research* 45, No. 4, Winter 1998, Alabama Agricultural Experiment Station, Auburn, AL available at <http://www.auburn.edu/aaes/communications/highlights/winter98/bollgard.html>
- Stone, G.D. (2004) Social constraints on crop biotechnology in developing countries. *AgBioForum* 7, 76–69. available at <http://www.agbioforum.org/v7n12/v7n12n14-stone.htm>
- Sunilkumar, G., Campbell, L.M., Puckhaber, L., Stipanovic, R.D., and Rathore, K.S. (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc. Nat. Acad. Sci. USA* 103, 18054–18059.
- Tabashnik, B.E., Gassmann, A.J., Crowder, D.W., and Carriere, Y. (2008) Insect resistance of Bt crops: evidence versus theory. *Nature Biotechnol.* 26, 199–202
- Tohidfar, M., Mohammadi, M., and Ghareyazie, B. (2005) Agrobacterium-mediated transformation of cotton (*Gossypium hirsutum*) using a heterologous bean chitinase gene. *Plant Cell, Tissue Organ Cult.* 83, 83–96
- Torres, J.B., and Ruberson, J.R. (2005) Canopy- and ground-dwelling predatory arthropods in commercial Bt and non-Bt cotton fields: patterns and mechanisms. *Environ. Entomol.* 34, 1242–1256.
- Torres, J.B., and Ruberson, J.R. (2008) Interactions of *Bacillus thuringiensis* CryI Ac toxin in genetically engineered cotton with predatory hemipterans. *Transgenic Res.* 17, 345–354
- USEPA (1998) U.S. Environmental Protection Agency, Pesticide Registrations, Federal Register May 13, 1998, Vol 63, # 92 available at <http://www.epa.gov/fedrgstr/EPA-PEST/1998/May/Day13/p12639.htm>
- USDA.AMS (2006) Cotton varieties planted: 2006 crop. U. S. Department of Agriculture, Agricultural Marketing Service, Cotton Program, Aug 2006, available at http://www.ams.usda.gov/cottonrpts/MNXLS/mp_cn833.xls
- USDA ERS (2004) Characteristics and production costs, statistical bulletins, U.S. Department of Agriculture Economic Research Service, available at <http://www.ers.usda.gov/publications/sb974>
- USDA ERS (2008) Cotton: trade, U.S. Department of Agriculture Economic Research Service, available at <http://www.ers.usda.gov/Briefing/Cotton/trade.htm>
- Van Deynze, A.E., Sundstrom, F.J., Bradford, K.J. (2005) Pollen-mediated gene flow in California cotton depends on pollinator activity. *Crop Sci.* 45, 1565–1570.
- Van Rie, J. (2006) Genes with plant-optimized codons encoding *Xenorhabdus* bacterial insecticidal protein and their use in producing insect-resistant transgenic plants. *PCT Int. Appl.* As cited in CAN:22267
- Wang, Y.Q., Chen, D.J., Wang, D.M., Huang, Q.S., Yao, Z.P., Liu, F.J., Wei, X.W., Li, R.J., Zhang, Z.N., and Sun, Y.R. (2004) Over-expression of *Gastrodia* anti-fungal protein enhances *Verticillium* wilt resistance in colored cotton. *Plant Breed.* 123, 454–459.
- Whitehouse, M.E.A., Wilson, L.J., and Fitt, G.P. (2005) A comparison of arthropod communities in transgenic Bt and conventional cotton in Australia. *Environ. Entomol.* 34, 1224–1241.
- Whitehouse, M.E.A., Wilson, L.J., and Constable, G.A. (2007) Target and non-target effects on the invertebrate community of Vip cotton, a new insecticidal transgenic. *Aust. J. Agric. Res.* 58, 273–285
- Wu, K.M., and Guo, Y.Y. (2005) The evolution of cotton pest management practices in China. *Annu. Rev. Entomol.* 50, 31–52.

- Wu, J., Guo, H., Xiao, J. and Tian, Y. (2006) Transgenic cotton expressing *Amaranthus caudatus* agglutinin, confers enhanced resistance to aphids. *Plant Breed.* 125, 390–394
- Wu, J., Luo, X., Wang, Z., Tian, Y., Liang, A., and Sun, Y. (2008) Transgenic cotton expressing synthesized scorpion insect toxin AaHIT gene confers enhanced resistance to cotton bollworm (*Heliothis armigera*) larvae. *Biotechnol. Lett.* 30, 547–554
- Wu, K.M. (2007) Environmental impact and risk management strategies of Bt cotton commercialization in China. *Chinese J. Agric. Biotechnol.* 4, 93–97
- Xu, B., Wang, Y., Liu, X., Yuan, F., Su, F., Chen, Y., Wu, Y., and Zhang, Q (2006) Effects of CryI Ac and secondary metabolites in Bt transgenic cottonseed on *Lycoriella pleuroti* Yang et Zhang (Diptera: Sciaridae). *Environ. Entomol.* 35, 807–810.
- Xue, D. (2002) A summary of research on the environmental impacts of Bt cotton in China. Nanjing Institute of Environmental Sciences, State Environmental Protection Administration of China, Greenpeace., available at http://archive.greenpeace.org/geneng/reports/env_impact_eng.pdf
- Yu, L., Berry, R.D., and Croft, B.A. (1997) Effects of *Bacillus thuringiensis* toxins in transgenic cotton and potato on *Folsomia candida* (Collembola; Isotomidae) and *Oppia nitens* (Acari: Oribatidae). *J. Econ. Entomol.* 90, 113–118.

Toward Characterizing the Spectrum of Diversity in the *Gossypium* Genus

Andrew H. Paterson

Abstract Sequencing of the human genome has driven the development of a host of technologies that have rendered DNA as the raw material for revealing and understanding biodiversity, integrating rich histories of phenotypic, genetic and evolutionary information with their molecular basis. The exciting day is foreseeable when we have most if not all of the *Gossypium* genomes sequenced. How will we then convert the sequence into knowledge, and into application? What will be the path? What will be the cost?

1 Biodiversity, Organic Evolution, and Agricultural Sustainability

A recurring theme throughout many chapters of this book has been that the forty-five diploid and five tetraploid species of the genus *Gossypium*, now distributed over six continents, offer a rich source of diversity useful in many ways toward deeper understanding of organic evolution, and which also provide intrinsic low-cost genetic solutions to a host of needs of the producers, processors, and consumers of the world's leading natural textile fiber. Of the broad spectrum of morphological and physiological diversity available in the genus (Wendel et al, Chapter 1), only a tiny fraction is reflected in the elite gene pools of *G. hirsutum* (Lubbers and Chee, Chapter 2), *G. barbadense* (Percy et al., Chapter 3), *G. arboreum* or *G. herbaceum* (Kulkarni et al, Chapter 4). The breadth of diversity in the *Gossypium* genus includes many valuable traits in the secondary and tertiary gene pools that might also be incorporated into the primary gene pool with a minimum of linkage drag (Chapters 2-4, 11-14) and a minimum of public objection (Davis et al, Chapter 19) if we could identify the responsible genes.

A.H. Paterson (✉)
University of Georgia, Plant Genome Mapping Laboratory
e-mail: paterson@uga.edu

This is not to downplay the importance of the variation remaining in the elite gene pools – although limited to a tiny fraction of the diversity in the genus, even relatively subtle morphological variations that adapt cotton to particular production regions or environmental hazards (Lubbers et al, Chapter 2) can be of major economic importance. The continually changing spectrum of disease and insect pests (and races therein) necessitate ongoing characterization and deployment of new host plant resistance genes and gene combinations (Wright et al, Chapter 13). Ever-growing urban demand offers growing competition for fresh water resources that are an important input to high-yield cotton production in many regions, and motivate more intensive searches for genes and mechanisms that permit the cotton plant to make a crop with less water, as well as to endure periods of water deficit (Saranga et al, Chapter 14).

Like many taxa that primitive peoples domesticated, the *Gossypium* genus is also a botanical model for shedding light on particular aspects of plant biology. Few if any other plants produce unicellular seed-borne epidermal fibers of 1-2" in length (Chee et al, Chapter 12). As such, cotton provides a novel system in which to better understand the genetic and biochemical control of cell elongation as well as a potential single-celled genomics system. The fact that these fibers are cloaked in some of the purest cellulose known in nature (Haigler et al, Chapter 16), offers the possibility to gain new insight into the biosynthesis of this most abundant of organic molecules, which is of increasing importance as a repository for solar energy before processing into ethanol or other biofuels convenient for our use.

2 Sequencing the *Gossypium* Genomes

Sequencing of the human genome has driven the development of a host of technologies that have rendered DNA as the raw material for revealing and understanding biodiversity. Once a specialized tool only applicable to small numbers of high value genes, DNA sequencing is now routinely performed in massive scale with 'Moore's Law'-like progress in throughput and cost-efficiency. Angiosperm (flowering plant) genomes recently comprised ~13.5% of Genbank sequence data: if this fraction is maintained - and if we sustain the 60% per year average increase in sequence information that has been realized since the 1980s – then the sequencing of the 200 most important domesticated plants might be largely completed by 2020 (Paterson 2006). Emerging technologies offer the means to maintain or perhaps even accelerate the increase in sequence information (Margulies et al, 2005; Shendure et al, 2005). This will eventually reduce dramatically the complexity of the decision to sequence a genome, making it possible to sequence multiple genomes routinely in the context of everyday research programs.

Cotton is unusual, although not unique, in that we will need to sequence not only cultivated (tetraploid) genotypes but their diploid progenitors, to understand how tetraploid cottons have come to 'transgress' the productivity and quality of their progenitors. In numerous ways from gene expression patterns

(Adams and Wendel, Chapter 17) to quantitative variations in productivity and quality (Chee et al, Chapter 12), the tetraploid species that dominate world cotton commerce have evolved emergent properties that substantially exceed the summed properties of their diploid progenitors.

Sequences from the A and D genome diploid species likely to most closely resemble the progenitors of tetraploid (AD) cotton, are obvious and high priorities for genome sequencing. Better understanding of gene repertoire and function in these species, in particular revealing differences in gene content and expression patterns across the ploidy levels, could prove invaluable in gaining insight into polyploid genome evolution. The high degree of conservation of gene order and sequence between diploids and tetraploids suggests that the vast majority of data from diploids will extrapolate directly to tetraploids. Accordingly, the worldwide cotton community has repeatedly endorsed a strategy focused on first obtaining a high-quality whole-genome sequence of the smallest *Gossypium* genome (D), toward the goal of also obtaining high-quality sequences for the A and AD tetraploid genomes (Paterson 2007; Z. Jeffrey Chen et al, 2007). Some technical questions remain about the details of the best strategies by which to sequence these genomes (Lin and Paterson, Chapter 6) that will be answered in the next year or two by pilot studies in progress.

Sequencing of representatives from each of the 8 genome types, will be important to molecular dissection of numerous evolutionary patterns and biological phenomena, including the genomic and morphological diversity that has permitted species within the genus to adapt to a wide range of ecosystems in warmer, arid regions of the world. If indeed the sequencing of the 200 most important domesticated plants takes 14 years, with continuing increases in sequence information of 60% per year one can envision that a few years later it is reasonable to have sequenced several species per crop lineage? Once gold-standard sequences are established for key species in a clade, further sequencing will be able to rely more and more heavily on rapid and dramatically lower-cost resequencing approaches that assume (quite safely in much if not all of the *Gossypium* genus) a high degree of colinearity. Finally, the scientific community has tended to underestimate the rate at which sequence information will expand – if the conservatism of scientists is once again outstripped by advances in technology, then we may realize the benefits of multiple *Gossypium* sequences sooner than we expect. Certainly within the next 20 years, and very probably sooner, it is a reasonable expectation that we may have workable whole-genome sequences for multiple members of most (if not all) of the *Gossypium* genome types.

3 Linking Sequence to Function

Once sequenced, how will the cotton genes and their variant alleles be linked to their phenotypic effects?

We will convert some cotton sequence to information by identifying similarities to other well-studied genomes. The relatively close relationship of cotton

and *Arabidopsis*, detailed genetic map for cotton, and potential importance of using functional genomic information and tools from *Arabidopsis* to aid in dissecting economically-important pathways in cotton make this system an excellent case study for exploring comparisons of gene order among divergent taxonomic families (Rong and Paterson, Chapter 18). Research into the genetic control of cotton fiber development may benefit from progress in understanding the growth and development of hair-bearing epidermal cells (trichomes) in *Arabidopsis*. Indeed, *Gossypium* and *Arabidopsis* are thought to have shared common ancestry about 83-86 million years ago (Benton 1993), and in view of this relatively close relationship cotton may be the best crop outside of the Brassicales in which to employ 'translational genomics' from *Arabidopsis*.

About 60% of cotton genes had one or more unambiguous homologs in the *Arabidopsis* genome, and about half of the cotton and *Arabidopsis* transcripts show correspondence in gene arrangements (Rong et al, 2005), suggesting that the well-annotated *Arabidopsis* genome sequence and its gene set, the best-studied of any plant, will provide considerable guidance in deducing the structure and function of those cotton genes that are relatively conserved.

To understand and manipulate the features that make cotton unique will require a host of enabling tools, technologies, and resources; in particular targeting portions of the sequence that are substantially different from those of other organisms. In that the basic gene set for angiosperms has largely been revealed by sequencing of several botanical models, a natural priority in sequencing cotton will be to reveal genes related to its unique features. There are few if any other examples of seedborne epidermal plant cells that reach 1-2" or more in length and are nearly pure cellulose. How will we recognize the genes that confer these features, and how will we determine how they work?

Computational approaches may help to identify some of the genes that are likely to contribute to the unique features of cotton. Rapid gene evolution may be due to a lack of structural or functional constraint, or to strong positive selection for functional divergence. Established statistical approaches allow one to computationally identify candidates that distinguish clearly between these possibilities (Yang 1997; Nielsen and Yang 1998; Yang 1998; Yang et al. 2000a). Genes under strong positive selection are an important complement to the highly conserved, functionally important genes amenable to comparative genomics. For example, rapidly evolving genes in *Drosophila*, mammals, and several other species are vital to reproductive success, cell-cell recognition, and cellular response to pathogens (e.g., (Yang et al. 2000b; Swanson et al. 2001a; Swanson et al. 2001b)).

However, recognition of genes that appear to have evolved rapidly will not by itself reveal their functions. More generally, there is every reason to expect that many cotton genes may have different (or at least partly different) functions than *Arabidopsis* genes with similar sequences. Even 'diploid' cottons are actually paleopolyploids, having incurred a large-scale (presumably whole-genome) duplication since their divergence from *Arabidopsis*. As a consequence

of genome duplication(s) and associated gene loss, gene linkage relationships in cotton are often different than in *Arabidopsis* (Rong et al. 2005). Further, there is every reason to anticipate that the functions of some genes have been subdivided [*subfunctionalized* – (Lynch and Force 2000)] between duplicated *Gossypium* copies, while other duplicated copies may have evolved completely new functions (neofunctionalization) that do not exist in *Arabidopsis* or other outgroups. This phenomenon happens very rapidly in synthetic tetraploids, clearly also happened in the evolution of natural tetraploids (Adams and Wendel, Chapter 17), and must be considered likely to have also happened following a probable genome duplication in the ancestors of modern diploid cottons perhaps 20 million years ago (Rong et al. 2004; Rong et al. 2005). Indeed, *Gossypium* is a singularly-attractive genus in which to shed new light on the morphological and evolutionary consequences of polyploidy, which is proving to be one of the most important forces shaping angiosperm genomes (Paterson 2008).

Finally, *Arabidopsis* itself is not nearly so ‘simple’ a genome as was once thought, having been through at least one whole-genome duplication itself since its divergence from cotton – thus, both *Arabidopsis* genes AND *Gossypium* genes may have changed function and tissue-specificity since their divergence from common ancestral genes. Ongoing sequencing of additional angiosperm genomes has suggested that papaya (Ming et al. 2008) and grape (Jaillon et al. 2007) may prove to be better models than *Arabidopsis* for deducing ancestral and derived (modern) functions of groups of related genes.

In partial summary, to understand and manipulate the features that make cotton unique will require new enabling tools, technologies, and resources. A few particularly-high priorities among these are likely to include (in random order):

1. Large-scale expression profiling of the full set of cotton genes (indeed, preferably the entire genome) across a comprehensive sampling of *Gossypium* species, tissues, organs and developmental states, using a common platform such as has been used in other taxa (Persson et al. 2005), to permit deductions about gene function based on coordinated expression patterns. While appreciable progress has already been made (Chapters 7, 16), whole-genome sequences are likely to reveal many genes of high importance but which are expressed too briefly and/or in too few cells to have been identified based on current information.
2. Large-scale sampling of patterns of between-species divergence and within-species diversity in the sequences of the full set of cotton genes (indeed, preferably the entire genome), providing the means to distinguish among genes that show evolutionary patterns such as:
 - a. Divergence to novel function in a particular clade (for example, the A-genome diploids), followed by purifying selection within that clade suggesting that the new function is under strong selection;

- b. Divergence to new function in a clade, with continuing positive selection within the clade such as might be expected in the ongoing ‘arms war’ between plants and their pests;
 - c. Conservative evolution across otherwise divergent clades, suggesting that the ancestral function is broadly adaptive and under purifying selection.
3. Comprehensive mutant resources. Strategies for *Gossypium* functional genomics need to anticipate that many genes may be implicated in crop improvement by association genetics approaches that would benefit from functional validation. Comprehensive mutant populations (Auld et al, Chapter 9), using established techniques (McCallum et al. 2000; Till et al. 2003; Slade et al. 2005; Comai and Henikoff 2006) that are likely to become much faster and less costly to characterize by using future-generation resequencing technologies, can provide a means by which functional analysis of *Gossypium* genes can be carefully-targeted to complement and supplement more extensive resources for *Arabidopsis* and other botanical models. This approach will provide for both the study of genes/gene families that are less tractable in other plants, and also for targeting functional analyses to specific genes implicated in key cotton traits by association genetics or other approaches. Such resources are ideally needed for each of the two cultivated tetraploids (to permit study of duplicated gene fates during all-important adaptation to the polyploid state) and each of the diploid genome types, with priority placed on the progenitor A and D genomes that contributed to the tetraploid.

What will it all cost?

In the crown botanical model *Arabidopsis thaliana*, for which the essentially-complete sequence was published in 2000 (Initiative 2000), the US National Science Foundation (www.nsf.gov) has spearheaded an effort to determine the functions of all *Arabidopsis* genes by 2010. This program has yielded an enormous body of knowledge relevant not only to *Arabidopsis* but to cotton and many other major crops, at a cost approaching \$200 million — and the job is not finished. While we will learn the functions of some cotton genes by analogy to those of *Arabidopsis*, as noted above many genes will remain to be characterized in cotton directly. Moreover, cotton has the further disadvantages of nearly 20 times as much DNA, a much larger plant body that clearly cannot complete its life cycle in a test tube(!), and the necessity of completing its much longer life cycle in order to see impacts of specific genes on the trait of primary importance (seedborne fibers). We must assume that to truly translate the cotton sequences into knowledge that can be applied on the farm, the cost may approach that expended for *Arabidopsis*. This is in no way a criticism of public investments in *Arabidopsis* — had *Arabidopsis* not gone first the cost of cotton functional genomics would have been many times higher.

Cotton, of course, has the advantage that it is the raw material for a number of high value industries, and that investments might be rewarded with lower production costs, more efficient processing, and new or improved products.

The perception of a decade ago by many life sciences ventures that DNA sequence alone might lead to a dominant intellectual property position has largely been replaced by an understanding that the pipeline from sequence to product is quite long. Many of the interim steps involve the development and application of enabling tools that individually are of little proprietary value but are nonetheless essential links in the chain between DNA sequence and economic benefit. An intuitively attractive concept is that the *Gossypium* industries and scientists band together, forming a public-private partnership to develop a full toolbox of widely-embraced enabling tools (and associated computational resources) that permit a rapid accumulation of information on common platforms, while providing a mechanism by which private partners can reserve intellectual property rights on specific discoveries resulting from the use of these tools. Of course, "... the devil is in the details."

In conclusion, one could hardly envision a more exciting time to be in the field of genetic analysis of any organism, and cotton is no exception. The hereditary molecules are being laid bare, their contents exposed by computational tools that are increasingly powerful and accessible. As a graduate student in plant breeding in the 1980s, walking the fields and taking notes, I never envisioned that the genomes of our major crops would be fully sequenced during the duration of my own career ... but it now seems a near-certainty. Indeed, we may see the day when our capacity to sequence genomes outstrips our ability to analyze their contents. Nonetheless, the mysteries remain – with their dissection, understanding, and application providing job security to geneticists for years to come.

References

- Benton MJ (1993) The fossil record 2. Chapman and Hall. New York
- Comai L, Henikoff S (2006) TILLING: practical single-nucleotide mutation discovery. *Plant J* 45:684–694
- Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Jaillon O, Aury JM, al. e (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* advanced online publication: doi:10.1038/nature06148
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473
- Margulies M, Egholm M, Altman WE, Attiva S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GPI, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380

- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology* 123:439–442
- Ming R, Hou S, Feng Y, Yu QY, Dionne-Laporte A, Saw J, Senin P, Wang W, Salzberg SL, Tang H, Lyons E, Rice D, Riley M, Skelton R, Murray J, Chen C, Eustice M, Tong E, Albert H, Paull RE, Wang ML, Zhu Y, Schatz M, Nagarajan N, Agbayani R, Guan P, Blas A, Wang J, Na JK, Michael T, Shikrov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Suzuki J, Tripathi S, Neupane K, Wei H, Singh R, Irikura B, Jiang N, Zhang W, Wall K, Presting G, Gschwend A, Li Y, Windsor AJ, Navajas-Perez R, Torres MJ, Feltus FA, Porter CH, Paidi M, Luo MC, Liu L, Christopher D, Moore PH, Sugimura T, dePamphilis C, Jiang J, Schuler M, Mitchell-Olds T, Shippen D, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–997
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936
- Paterson AH (2006) Leafing through the genomes of our major crop plants: strategies for capturing unique information. *Nature Reviews Genetics* 7:174–184
- Paterson AH (2007) Sequencing the cotton genomes. World Cotton Research Conference. International Cotton Advisory Committee. Lubbock TX
- Paterson AH (2008) Paleopolyploidy and its Impact on the Structure and Function of Modern Plant Genomes. *Genome Dynamics* 4:1–12
- Persson S, Wei HR, Milne J, Page GP, Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences of the United States of America* 102:8633–8638
- Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Research* 15:1198–1210
- Rong JK, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding XL, Garza JJ, Marler BS, Park CH, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao XP, Zhu LH, Paterson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166:389–417
- Shendure J, Porreca GJ, Reppas NB, Lin XX, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–1732
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* 23:75–81
- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001a) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 98:7375–7379
- Swanson WJ, Zhang ZH, Wolfner MF, Aquadro CF (2001b) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America* 98:2509–2514
- Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research* 13:524–530
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573
- Yang Z, Nielsen R, Goldman N, Krabbe Pedersen A (2000a) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155

- Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13:555–556
- Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000b) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Z. Jeffrey Chen. Brian E. Scheffler. Elizabeth Dennis. Barbara Triplett. Tianzhen Zhang. Xiaoya Chen. David M. Stelly. Pablo D. Rabinowicz. Christopher Town. Tony Arioli. Curt Brubaker. Roy Cantrell. Jean-Marc Lacape. Mauricio Ulloa. Peng Chee. Alan R. Gingle. Candace H. Haigler. Richard Percy. Sukumar Saha. Thea Wilkins. Robert J. Wright. Allen Van Deynze. Yuxian Zhu. Shuxun Yu. Wangzhen Guo. Ibrokhim Abdurakhmonov. Ishwarappa Katageri. Mehboob-ur-Rahman. Yusuf Zafar. John Z. Yu. Russell J. Kohel. Jonathan Wendel. Paterson AH (2007) Toward sequencing cotton (*Gossypium*) genomes. . *Plant Physiology* 145:1303–1310.

Index

A

- Aarts, M.G., 220
 - Abiotic stress resistance
 - gene expression studies of
 - microarray technology and, 344–345
 - physical map, 344
 - tetraploid species GH and GB
 - arid conditions, 343
 - GhNHX1, 346
 - heat and drought stress, 341, 342
 - irrigation regimes, 343
 - stomatal conductance and transpiration, 342
 - transgenic approach to improve, 345–346
 - Acalas, 32
 - Ac/Ds transposon system
 - LTR-retrotransposons and, 219
 - TPase activity and CaMV35S promoter, 217
 - Acharya, S., 75
 - ACO genes, 177
 - Adams, K. L., 426
 - AdhA BAC regions, analysis of, 421
 - Adragna, H., 298
 - AD tetraploid species, 83
 - AFLP, *see* Amplified fragment length polymorphism
 - AFLP-RGA, 118
 - African-Arabian species of *Gossypium*, 8
 - genetic diversity in, 9
 - taxonomic section, 9
 - A-genome species
 - domesticated, 13
 - seed dispersal in, 15
 - Agrobacterium tumefaciens*, 188
 - Agronomically important genes, physical mapping of, 272
 - Al-Khatib, K., 471
 - Allopolyploid cottons
 - diversification of, 10–11
 - DNA content of, 152–153
 - duplicates, 13
 - genomes (A and D), 13
 - phylogenetic history of, 420
 - pleistocene origin for, 10
 - Alternative oxidase, 197
 - American Pima cottons
 - gene pool, sources of, 59
 - Pima S-2, 59
 - Pima S-3, 60
 - productivity and fiber quality, 60
 - progenitors of, 58, 59
 - SxP, 58
 - Amplified fragment length polymorphism and cleaved AFLP
 - genetic similarity, 118
 - restriction enzymes, 117
 - in combination with degenerate primers, 118
 - linkage and QTL mapping using, 117
 - mechaism of operation of, 116
 - methylation polymorphism pattern studies by, 118–119
 - phylogenetic and diversity studies using, 116
- Annuals, 37
 - AOX, *see* Alternative oxidase
 - Applequist, W. L., 37
 - Applied breeding
 - breeding concerns in, 42
 - Mepiquat-based PGRs use in, 42–43
 - selectivity in weed control in, 43
 - Arabidopsis* and cotton genomes,
 - comparative genomics of
 - evolutionary context of, 438
 - fiber QTLs, 442–443
 - MYB gene expression, 441–442

- Arabidopsis* and cotton genomes,
 comparative genomics of (*cont.*)
 synteny between, 438–439, 444
 trichome development, 440–441
- Arabidopsis thaliana*
 cell-wall-associated genes of
 CESA genes, 404
 COBRA-LIKE4 gene and *POM1*/
 AtCTL1 gene, 405
 co-expression with secondary wall
 CESA genes, 403
 microarray results for cotton
 homologs of, 402
 coding sequence homology between
 cotton, 165
 BLASTX hit, 167
 in phylogenetic context, 169
DREB1A overexpression in, 346
 genome of
 cotton-*Arabidopsis* synteny, 444
 and cotton genomes, comparative
 genomics of, *see Arabidopsis* and
 cotton genomes, comparative
 genomics of
 duplication in, 438
GF14l gene, overexpression of, 346
 homologs of up-regulated genes, 397
 Initiative 2000, 488
 leaf trichomes and cotton seed trichomes,
 similarities between fiber initiation
 of, 174–176
 Na^+/H^+ antiporter overexpression in, 346
 pathogen defense genes and BRR QTLs,
 relationship between, 324
 proteome and cotton fiber gene
 expression, 393–395
 reverse genetics via TILLING, 214
 secondary wall deposition, genes related to
 carbon supply to cellulose synthetic
 enzymes, 407–408
 CESA isoforms, 404
 COBRA gene, 404–405
 cotton homolog of *At2g37090*,
 406–407
 CSCs, 408
 GhCESA4 gene, 403
 G.h.fbr-sw sequence, 410
 homologous genes to *Arabidopsis*
 FLA11 and FLA12, 409
 KNAT transcription factors, 405
 KORRIGAN protein, 408
 POM1/AtCTL1 mutation, 405
 RHD3 and KOBITO, 409
 SPIRALI-LIKE genes, 406
 seeds, γ -TMT overexpression in, 373
 SNPs in, 125
 T-DNA insertions in, 216
 transposition of *Tnt1* in, 219
 transposon tagging using *Ac/Ds*
 elements in
 En/I system, 218
 LTR-retrotransposons, 219
 stable TPase source, 217
 TIRs, 218
 trichome development in, 440–442
 Array Probe Portal, 239
 Ashmouni, 56
 Australian cottons, 7
 C-, G-, and K-genome groups in, 8
 insect community in canopies of Bt and
 non-Bt cotton on, 463
 Xcm resistance in, 322
 Autonomous mini-chromosomes, cotton
 transformation using, 195–196
- B**
 Backcrossing, 87
 Backcross-QTL schema, 45
 Backe, E. E., 287
 BAC libraries, 344
 FISH mapping using, 273
 of *G. hirsutum*, 150–151
 probe-BAC hybridization data for, 236
 search criteria, 232, 244
 BAC sequencing studies, polyploid cotton
 genome evolution, 421
 Bacterial artificial chromosome libraries, *see*
 BAC libraries
 Bacterial blight resistance gene resistance
 B₂, B₃, b₆, and B₁₂ genes, 320–321
 B_{6m} intensified resistance, 321
 Bambawale, O. M., 468
 Bayley, C. C., 200
 Beal, J. M., 261
 Beasley, C. A., 190
 Beasley, J. O., 259, 262
 Bengal Desi, 73–74
 ‘bengalense’ (G 135-49) race, 92
 Bergey, D. R., 270
 Bernardo, R., 305
 Berry, R. D., 458
 Bhosle, B. B., 468
 Biggs, R. W., 463
 BioMOBY services, 229
 Biotech cotton, *see* Bt cotton

- Black root rot, 323
- Blast2GO, 166
- BLASTing sequences, 144
- BLASTX, 166
- Bolek, Y., 327
- Boll weevil invasion, 31–32
- Bourland, F. M., 73
- Bowman, D. T., 38–39
- Breeding, 41
 - of American Pima, 60
 - of *G. barbadense*
 - goals of, 62
 - heterotic responses in, 63
 - for improvement of oil quality using
 - spontaneous mutation, 358
 - molecular, 330–331
 - of Sea Island cottons, 55–56
- Brown, M. S., 263
- BRR, *see* Black root rot
- Brubaker, C. L., 27–28, 73
- Bt cotton, 456
 - broad effects of
 - Bt toxin, 458
 - lepidopterans and syngenta, 457
 - life cycle of predator, 458
 - development and adoption of, 454–455
 - impact on pests and predators
 - on African continent, 469–470
 - biodiversity, 462–463
 - in Brazil, 469
 - in China, 465–467
 - CpTI (cowpea trypsin inhibitor), 465
 - in India, 467–469
 - insect community, Australian cotton farms, 463
 - in Mexico, 469
 - plant bugs, 461
 - predaceous arthropods, 459–460
 - relay-cropping, 463–464
 - potential for benefits from
 - insect resistance, 453
 - resistant to broad-spectrum herbicides, 454
 - without transgenes, 455
- Burke, J., 195
- Burmanicum, 73
- C**
- Calhoun, D. S., 38–39
- Canola quality oil, 358
 - stearic acid level in, 367
- Cantrell, R. G., 455
- Carlson, S. R., 195
- Carriere, Y., 463
- CDD, *see* Cotton Diversity Database
- cDNA microarrays
 - for cotton, *see* Cotton cDNA microarray
 - secondary wall deposition, genes associated with, 403
- CEGC, *see* Comparative Evolutionary Genomics of Cotton
- CEL I, single strand-specific nuclease, 214–215
- Cellulose
 - biogenesis of, *see* Cellulose biogenesis
 - definition of, 385
 - impact on fiber physical properties, 389
 - fiber strength, 391
 - secondary wall thickness and percent crystallinity, 390
 - molecular properties of, 390
- Cellulose biogenesis
 - CESAs and CSC role in, 387
 - cotton fiber
 - as experimental tool understanding, 386
 - primary walls, 388
 - secondary wall deposition, control of, 388–389
 - localization and orientation of, 388
- Cellulose fibrils, 385
- Cellulose nanofibril orientation, 386
- Cellulose synthase (CESA), 387
 - interaction via N-terminal zinc-binding domain, 389
 - isoforms, up-regulation of, 404
- Cellulose synthesis complex, 387
- Central Institute of Cotton Research (CICR), 77
- ‘cernum’ race
 - boll size, 89
 - immune reaction to grey mildew, 90, 92
- Chee, P., 301
- Chee, P. W., 40
- Chemical mutagenesis
 - ethylmethane sulfonate treatment
 - fiber length, 213
 - seed viability, 211
 - sodium azide, 210
- Chemical mutagens, 214
- Chen, X. X., 465
- Chlan, C. A., 194
- Chromosome morphology, 7
- Chua, N. H., 200
- Cienfuegosia*, 4
- CINA 36, 89

- Classical cytogenetic studies
 - breeding programs
 - bispecific introgression pathway, 269
 - interspecific breeding programmes, 263–264
 - meiotic analysis, 264–267
 - tri-specific pathway, 268
 - cotton responses to abiotic stress
 - genetic variability for, 341–342
 - GH and GB, 342
 - diploid *Gossypium*
 - chromosome morphology, 260
 - genomic classification, 259
 - paleopolyploid origin, 261
 - principles of, 258–259
 - tetraploid *Gossypium*
 - allotetraploids, 261
 - chromosome length, 262
 - translocations, 262–263
- Class II (DNA) transposable elements, 148
- Class II (DNA) transposons, superfamilies of, 145
- Class I transposons, 145
- Clearfield technology, 455
- Cleaved AFLP
 - and AFLP
 - genetic similarity, 118
 - restriction enzymes, 117
- COBRA* gene, 404–405
- Coetzer, E., 471
- Colchicine, 211
- Comparative Evolutionary Genomics of Cotton, 237
- Comparative genomics
 - applications of, 432
 - cotton and *Arabidopsis* genomes
 - evolutionary context of, 438
 - fiber QTLs, 442–443
 - MYB gene expression, 441–442
 - synteny between, 438–439, 444
 - trichome development, 440–441
 - foundation of, 433
- Corolla colors, 7
- Cotton
 - abiotic stress resistance of, *see* Abiotic stress resistance
 - as botanical model, 484
 - coding sequence homology between *Arabidopsis*, 165
 - disease resistance in, *see* Disease resistance
 - diversification of, 434
 - DNA reassociation kinetics studies of, 142
 - domestication of, *see* Domestication
 - general commercial breeding approaches for, 42
 - general transcriptome of, 165
 - genome composition characterization, 102
 - Class I transposable elements, 148
 - diploid progenitor, 153
 - by DNA reassociation kinetics, 143–144
 - repetitive sequences, 144–146
 - genome sequencing of
 - close relationship with *Arabidopsis*, 486
 - computational approaches for, 486
 - A and D genome diploid species, 485
 - emerging technologies for, 484
 - large-scale expression profiling, genes, 487
 - mutant populations, 488
 - patterns of between-species divergence and within-species diversity, 487–488
 - genome size variation in
 - BAC resources, 153
 - CesA1 region, 148
 - class I transposable elements role in, 147–148
 - indel numbers, 149
 - intron size, 148–149
 - repetitive DNA, 152
 - from *Gossypium* species, 23
 - importance in world economy, 24
 - improvement of
 - applied breeding, 42–43
 - breeding programs, 41
 - developmental breeding, 43–46
 - fiber quality, 40
 - QTL analysis, 46
 - origin of, 340–341
 - transformation of, *see* Cotton transformation
 - in vivo transformation of
 - mini-chromosomes, 195–196
 - naked DNA into axil placenta, 195
- Cotton cDNA microarray
 - long oligonucleotide probes, 171–172
 - short oligonucleotide probes, 172
- CottonDB (CDB), 234–235
- Cotton diseases
 - losses due to, 317
 - management (control) of
 - gene resistance, 318
 - resistant cultivars, 316
 - weed and insect pests, 316

- mechanisms of, 314
- pathogens, 315–316
- as tertiary problem, 313
- and US cotton production, 313, 316
- Cotton Diversity Database, 235
- Cotton EST assemblies, 162–163
- Cotton fiber, 35
 - developmental stages of, 173
 - genetic enhancement of, 209, 284
 - quality of, *see* Cotton fiber quality
 - secondary walls, 386
 - worldwide production of, 284
- Cotton fiber quality
 - definition of, 285
 - fiber properties
 - fiber elongation, 287
 - fiber length and length distribution, 286
 - fiber strength, 286–287
 - fineness and maturity, 287
 - genetic basis of, 432
 - genetic improvement of, 284
- Cotton fiber secondary wall transcriptome, analysis of
- cDNA library production and EST sequencing, 391–392
- cotton fiber gene expression *vs.* *Arabidopsis* proteome
 - genetic complexity during primary and secondary wall deposition, 393
 - G.h.fbr-sw sequences, 394
- for secondary wall deposition
 - Arabidopsis* homolog, 399–401
 - endogenous controls for qPCR analysis, 397–398
 - gene expression at 10 DPA, 400–401
 - genes expression at 20 DPA *vs.* 6 DPA and 10 DPA, 396
 - genes up-regulated at 24 DPA, 397
 - GhCTL1 expression, 398–399
- vs.* xylem, 401–403
- Cotton genetic maps, 122
 - genetic marker density, variations in, 149
- Cotton genus (*Gossypium* L.)
 - diversity and geographic distribution of major lineages of, 6–7
 - domesticated species, 4
 - evolutionary history of, 5
 - genetic enhancements of, 18–19
 - origin of, 4–5
 - phylogenetic relationships in
 - African F-genome clade, 13
 - allopolyploid cottons, 13
 - genealogical lineages, 11, 12
 - seed and trichome diversity in, 13–14
 - seed dispersal in, 15
 - speciation and diversification of
 - African-Asian species, 8–9
 - American diploid species, 9–11
 - Australian species, 7–8
 - chromosome morphology, 7
 - seed coverings, 7
 - species morphologies, 7
 - species-rich regions, 5
 - taxonomy of, 5
- Cotton leaf curl disease (CLCuD), 111
- “Cotton microsatellite database” (CMD), 120
- Cotton production
 - costs associated with, 453
 - genetic basis of drought and salt tolerance in, 443
 - high quality cotton, 454
 - in Israel and Peru, 61
 - supply and demand, 452
 - of USA, 29, 33, 43, 313
- Cotton seed
 - commercial, 198
 - epidermal layer of, 284
 - treatment with colchicine and gamma radiation, 211
 - yield of, 452
- Cotton seed oil, 198
 - fatty acid biosynthesis in, 361–362
 - fatty acid composition of
 - oleic acid, 356–357
 - polyunsaturated fatty acid, 356
 - research on modifying, 358
 - RNAi silencing of Fad2 and SAD-1 genes, 367–368
 - stearic acid, 357
 - genetic modification of
 - CPFAs, 369
 - endogenous gene expression, 360
 - gossypol reduction, 371–372
 - HO-CSO, 362–365
 - HS-CSO, 365–367
 - palmitic acid levels, 368–369
 - RNAi silencing of Fad2 and SAD-1 genes, 360–361, 367–368
 - unusual fatty acids production, 369–371
 - minor constituents of, improvement in
 - N-acylethanolamines, 373
 - phytosterols, 372
 - vitamin E, 372–373
 - nutritional and industrial value of, 355

- Cotton seed oil (*cont.*)
 prospects of increased utilization of, 357–358
 refined, 355
 uses of, 354
- Cotton TILLING Project, 215–216
- Cotton transformation
 by *Agrobacterium* infection
 cotton shoot apex, 194–195
 T-DNA, 193–194
 cotton seed, 198
 by micro-projectiles, 194
 plant gene expression, altering
 yield and stress responses, 197
 tGURTs, 200
 transgenics
 gene knock out population, 200
 resistant to herbicides, 196
 resistant to weed, 197
 for weed resistance, 197
- Cotton unigenes
 distribution with BLASTX hit, 166
 5 – 10 DPA group, 394
 between EST assemblies, 164–165
 frequency of gene ontology terms
 associated with, 167
- CPFAs, *see* Cylopropenoid fatty acids
- Crane, C. F., 271
- Cranell, C. F., 271–272
- Croft, B. A., 458
- Cronn, R., 37
- Cronn, R. C., 10, 423
- Crops productivity, effect of abiotic stress on
 in temperate and tropical areas, 337
 water deficiency, 338–339
- Crown gall disease, DNA involved in, 188
- CSCs, 408
- Culp, T. W., 37, 291
- Culpepper, A. S., 470
- Cultivated cottons
 genetic diversity in
 pre- and post- domestication
 interspecific gene flow, 16
 in series of early isozyme studies,
 17–18
 lint percentage of, 36
- Cultivated lint fiber, 15
- Cultivated tetraploid cotton
 biotic and abiotic stress tolerance of, 70
 BRR resistance in, 323
 interspecific backcross population
 of, 106
- Cylopropenoid fatty acids, 369
- Cytogenetics
 classical, *see* Classical cytogenetic studies
 definition of, 257
 molecular, *see* Molecular cytogenetics
- Czeschin, D. G., 271
- D**
- DAF assay, 109
- Day-neutrality, 44
- De Donato, M., 271–272
- DellaPenna, D., 373
- Demol, J., 82
- Dennehy, T. J., 463
- Dessau, D., 27
- Developmental breeding, 43
 chromosomal anomalies and, 45
 manipulations, 45–46
 natural crossing and selection, 44
 secondary gene pools, *G. barbadense*
 hybrids, 44
 tertiary gene pool, fertile hybrid, 46
- de Wet, J. M. J., 24–25
- DGAT activity and unusual fatty acid
 production, 371
- D-genome species
 endemics in, 9
 seed hairs in, 14–15
 sequence rate acceleration in, 423
- Dhandapani, A., 468
- Dhummad cotton, 75
- Diploid A-genome accessions, USA, 78
- Diploid cultivated cotton, 69
 A-genome, 70
 area of, 76–77
 biotechnological studies, genetic
 materials for, 93
 breeding approaches for
 boll size and fiber quality
 improvement, 88–90
 disease resistance and insect tolerance,
 90–93
 GMS and inter-specific hybridization
 for, 87
 inter-racial crosses and mass selection, 86
 seed cotton yield improvement, 87–88
- breeding of
 genetic improvement, in India, 83–84
 genetic variability, exploitable, 85–86
 hybrids, 84
 major research centers for, 84–85
 distribution in India
 Bengal Desi, 73–74

- Dhummad cotton, 75
- Mathio cotton, 74
- Wagad/Kalyan types, 75
- domestication and distribution of
 - G. arboreum*, 73
 - G. herbaceum*, 71–72
- genetic diversity
 - fiber properties and accessions, 81
 - morphological characters and karyological observations, 82
- germplasm resources
 - germplasm expedition, 79–81
 - global germplasm collection, 77–79
- global area of, 77
- K genome and D genome, 142
- morphological characters of, 70
- origin and evolution of, 71
- paleopolyploidy of
 - cytogenetic, EST, and structural genomic evidence for, 435–437
 - genetic diversity, 434
 - neopolyploidy, 434–435
- phylogenetic history of, 420
- regional diversity among, 81
- relationship with other genomes
 - A-genome species, 82–83
 - C- and G-genome, 83
- Disease resistance
 - bacterial blight pathogen
 - B₂ and B₆ gene, 321–322
 - B₁₂ locus, 322
 - D-subgenome, 320
 - gene resistance to *Xcm*, 320–321
 - R-genes, 322–323
 - BRR pathogen
 - PI 1415, 323
 - QTL regions, comparative analysis, 324
 - CLCuV disease, 326
 - factors influencing measures of, 319
 - Fusarium wilt, 324–325, 471
 - gene-for-gene models, 318
 - molecular dissection of, 319–320
 - reverse and comparative genetics of
 - ESTs and genome-wide expression, 329
 - resistance gene analogues, 329–330
 - speciation events and, 319
 - verticillium wilt
 - BAR gene, 472
 - impact on cotton production, 471
 - putative QTLs, 328–329
 - resistance to defoliating pathotype, 327–328
- Disease resistance (R) genes, 118
- Disease Triangle, 314
- DNA markers
 - associated with fiber properties, 113
 - associated with leaf hairiness, 114
 - chromosomal regions, 149
 - identification in cultivated cotton, 102
 - for physico-morphological traits, 113–114
- DNA methylation polymorphism patterns, 118–119
- DNA molecular marker technology, 39
- DNA reassociation kinetics, tetraploid cotton, 143–144
- Domestication
 - definition of, 26
 - of *G. hirsutum*
 - basal lineage of races of, 27
 - ‘marie-galante,’ 28
 - mocó cotton, 33
 - ‘punctatum’ and ‘latifolium,’ 28
 - genetic change caused by
 - in breeding and introgression, 40
 - coefficient of parentage, 39
 - diversity structure, upland cotton cultivars, 38
 - methods to analyze, 38–39
 - pattern of
 - fiber length, 37
 - flattening and convolution, 36
 - hard seed coats, 36
 - seed dispersal, 35
 - phenotypic consequences of, 34–35
 - of upland cotton
 - eastern big boll type, 30
 - genetics of, 29
 - long limb type, 31
 - semi-cluster type, 30
- Domestication syndrome of cotton, 37
- Donegan, K., 458
- Dooryard cottons, 34
- DREB1A, stress induced transcription factor, 346
- DRE-mediated gene expression, transcriptional repressor for, 346
- Duggar, J. F., 29
- E**
 - Early type cultivars, 31
 - Eastern big boll type, 30
 - Egyptian cottons
 - acclimatizing, 58
 - concern in early improvement efforts of, 57

- Egyptian cottons (*cont.*)
 fiber quality of, 57–58
 Jumel and Ashmouni, 56
 Mit Afifi, Sakel, and Giza 7, 58
 selection and improvement, 57
- Endrizzi, J. D., 10
- En/I system, 218
- En/Spm-like sequences, 145
- Entrez search tool, 236
- EST assembly
 browse and search capabilities of, 164
 of cotton
 number of unigenes in, 164–165
 programs for creating, 162–163
 cotton unigenes between, 164–165
 creation by ESTInformatics, 164
 quality of, 163
- ESTInformatics assembler
 Blast results of, 166
 TGICL tool and, 164
- EST unigene, identification of, 168
- Expressed sequence tag (EST)
 assembly of, *see* EST assembly
 cDNA transcripts, 158, 329
 collection and contributors, 158–161
 dynamic nature of, 164
 intra-genomic genome duplication, 437
 of suppressive subtractive hybridization
 library, 178
- Ex situ germplasm collections, 64–65
- F**
- FAD2 dimers, 363
- Feng, H. Q., 465
- Fiber length and elongation phase, 15
- Fiber quality improvement, 88–89
- Fiber quality properties
 genetic variation of, 289–291
 heritability studies of, 288–289
 improvement of, 291–292
 molecular genetics of, quantitative
 variation, 292
See also Cotton fiber quality
- Fiber transcriptome, 173
 cell elongation and primary cell wall
 formation
 ACO genes role in, 177
 water-mediated cell turgor, 176
 fiber initiation
 of cotton seed trichomes and
 Arabidopsis leaf trichomes,
 174–176
- MYB genes role in, 175
 putative transcription factors role
 in, 175
 regulatory genetic components of, 174
 transcription signal for, 174
 secondary cell wall formation and
 maturation, 178
- Finer, J. J., 194
- FISH
 in molecular cytogenetics, 270
 principle of, 269
- Fitt, G. P., 463
- Fluorescent in situ hybridization, *see* FISH
- French cotton germplasm collection, 78
- Fryxell, P. A., 11
- Fuzz fiber, 284, 440
- Fuzzy cotton seed, 353–354
- G**
- Gannaway, J. R., 40
- Gardner, A. P., 471
- Garro hill cotton, 89
- GeneChip[®] Cotton Genome Array, 172
- Gene ontology (GO)
 terms of
 ACO genes, 177
 Arabidopsis genes, 165–166
 associated with cotton unigenes, 167
- Genes Under Restriction Technology, 200
- Gene tagging
 by Ac/Ds transposon system, 217
 T-DNA for, 213
 by TEs, 216
 using RFLP, 107
 using SSRs, 123
- Genetically modified cotton seed oils, fatty
 acid composition of, 363
- Genetic/physical distance variation, 149
- Genome physical mapping, 273
- Genomic in situ hybridization
 applications of, 275
 labelled probe in, 274
- Germplasm pools, modern improved, *see*
 Modern improved germplasm
 pools, origins and development of
- Germplasm resources, diploid cultivated
 cotton
 AFLP analysis of, 116
 expedition of, 79–81
 global germplasm collection, 77–79
- G-genome species, seed dispersal in, 16
- Gharo cotton, 75

- GhBRI1* gene, base substitutions in, 215–216
- GhCESA4* gene, 403
- ghFAD2-1 expression and HO-CSO, 363–364
- G.h.fbr-sw sequence, 410
- GH_TMO ovules, EST analysis of, 170
- GISH, *see* Genomic in situ hybridization
- Giza 7, 56, 57
- GMS technique, 88
- Golovkin, M. V., 196
- Gossypieae*
- genera, 4
 - genome size and evolutionary relationship among, 142
 - phylogenetic relationships among, 111
- Gossypioides*, 5
- Gossypium anomalum*
- RAPD analysis of
 - fiber quality traits, 113
- Gossypium arboreum* L.
- base substitutions in *GhBRI1* gene, 215–216
 - 'bengalense,' 76
 - breeding approaches for
 - boll size and fiber quality improvement, 88–90
 - disease resistance and insect tolerance, 90–93
 - GMS and inter-specific hybridization for, 87
 - inter-racial crosses and mass selection, 86
 - seed cotton yield improvement, 87–88 - breeding of
 - genetic improvement, 83
 - genetic variability, exploitable, 85–86
 - hybrids, 84 - distribution of
 - in India, 73–75
 - indicum and burmanicum, 73 - domestication of, 4, 72
 - genetic diversity of, 9
 - fiber properties and accessions, 81
 - morphological characters and karyological observations, 82
 - RAPD analysis of, 111 - land races of, 77
 - long and narrow lobed leaves in, 70
 - molecular investigations of, 70
 - origin and evolution of, 71
 - progenies through introgression of *G. hirsutum* and
 - boll size, locule size, and halo length, 91
 - seed size, 92
 - yield and fiber quality characters, 90 - putative distribution of paleo-homoeologs in, 168–169
 - seed oil content in, 359
 - Vietnamese cotton collections, 79
- Gossypium barbadense*
- breeding of, 62–63
 - chemical mutagenesis in, 211
 - domestication of, 54
 - genetic variability in
 - ex situ germplasm collections, 64–65
 - habitat loss, 64 - species diversity of
 - allozyme investigation of, 63
 - geographic structure of, 54
 - taxonomic structure of, 53
- Gossypium bickii*, seed dispersal in, 15
- Gossypium* bioinformatics resources
- automated access to
 - Web services, 244, 245–246
 - workflows, 246–247 - data access features of, 233–234
 - data storage and accessibility, 231–232
 - genetic map and diversity data, 227
 - life science areas
 - crop science and structural genomics, 230
 - data types associated with, 230–231
 - resources within, 229 - navigational aids in, 228
 - search methods and strategies for
 - biochemical data, 243–244
 - genetic diversity, 238–241
 - genetic maps, 234–235, 237
 - genomic sequences and sequence tag sites, 235–236
 - phenotypic data range, 243
 - transcriptome and protein related data, 241–243 - use of ontologies by, 233
- Gossypium darwinii*
- domestication process, 4
 - genetic diversity in, 18
 - seed dispersal in, 15
- Gossypium* GenBank EST libraries, 159–162
- Gossypium gossypoides*, 9
- A-genome specific repetitive elements in, 146
- Gossypium herbaceum* L.
- acreage of, 75–76
 - breeding approaches for

- Gossypium herbaceum* L. (cont.)
 boll size and fiber quality improvement, 88–90
 disease resistance and insect tolerance, 90–93
 GMS and inter-specific hybridization for, 87
 inter-racial crosses and mass selection, 86
 seed cotton yield improvement, 87–88
 breeding of
 genetic improvement, 83
 genetic variability, exploitable, 85–86
 hybrids, 84
 distribution of
 in Africa and Asia, 72
 in India, 75
 northward spread, 71
 domestication of, 72
 domestication process, 4
 fuzzy mutant of, 93
 and *G. anomalum*, crosses between, 83
 genetic diversity
 fiber properties and accessions, 81
 morphological characters and karyological observations, 82
 germplasm of Iran, 78
 leaves in, 70
 molecular investigations of, 70
 origin and evolution of, 70
- Gossypium hirsutum*
 AFLP screen of MP diversity at CCGG sites, within, 119
 BAC library of
 male-sterile fertility restorer line 0-613-2R, 151
 ‘TM-1,’ 150–151
 categorization of, 34
 chemical mutagenesis in
 colchicine, 211
 ethylmethane sulfonate, 211–213
 sodium azide, 210
 classical cytogenetic studies of
 bispecific introgression pathway, 269
 interspecific combination, 268
 introgression, 268–269
 meiotic chromosome configurations, 267
 nonhomologous chromosomes, 262
 somatic chromosome length, 261–262
 translocation lines, 263
 cotton seed oil produced by, 355
 domestication process, 4
 evolutionary structure of, 24–26
 and *G. tomentosum*, AFLPs as diagnostic markers for, 117
 gene introgression, 89
 genetic diversity among, 18
 AFLP analysis of, 116–117
 RFLP analysis in, 103–107
 homoeologous genes in
 effects of abiotic stress conditions on, 425–426
 organ-specific expression biases of, 424–425
 introgression from *G. barbadense* into
 frequency of SNPs in, 126
 polymorphic AFLP-RGA fragments for, 118
 RAPD analysis of, 112
 SSR markers screened on, 120–122
 molecular cytogenetic studies of
 chromosome structural abnormalities, 272
 linkage groups, 273
 rDNA sites, 271
 primary wall deposition in, 388
 radiation mutagenesis in, 210
 reassociation kinetic analysis of, 143–144
 resistance against jassids in, 113
 RFLP genetic linkage map, 107
- Gossypium mustelinum*, genetic diversity in, 18
- Gossypium raimondii*
 BAC library of, 150
 whole genome sequencing of, 248
- Gossypium tomentosum*
 abiotic stress resistance of, 340–341
 and *G. hirsutum*, AFLPs as diagnostic markers for, 117
- Gossypium tomentosum*, genetic diversity in, 18
- Gossypium* transcriptome
 ancient duplications
 methods to identify, 168–169
 primary and secondary duplication events, 169
 characterization of
 direct sequencing of ESTs, 158, 162–165
 definition of, 157
 fiber quality, heritable differences in, 157
 gene sequences responsible for, 158
 homoeologous loci, 169–170
 microarray development for analysis of, *see* Microarray development,
 Gossypium transcriptome analysis

Gossypium trifurcatum, 9
 Gossypol
 reduction in cotton seed, 371–372
 research efforts, 199
 Gourani cotton, 75
Grandicalyx, 8
 Green, A., 198
 Green, C. C., 291
 Grey mildew disease, 90
 Guo, X. Y., 465
 GURT, *see* Genes Under Restriction Technology

H
 Hanson, R. E., 421
 Hard seed coats, 36
 Harlan, J. R., 24–25
 Harland, S. C., 54
 Harrell, D. C., 37
 Hasenkampf, C. A., 263
 hAT-like sequences, 145
 Hau, B., 27
 Herbicide resistance, 470–471
 Heterosis breeding, 89
 High-laurate canola, 368
 High-oleic cotton seed oil
 applications of, 365, 373–374
 development of
 expression of antisense/hairpin structure, 364
 ghFAD2-1 expression, down regulation of, 363
 transgene insertion, 364
 oxidative stability of, 364
 High stearic cotton seed oil
 applications of, 366
 melting point of, 366
 SAD gene, 366–367
 HO-CSO, *see* High-oleic cottonseed oil
 Hodges, T. K., 200
 Homoeologous gene
 effects of abiotic stress conditions on, 425–426
 organ-specific expression biases of, 424–425
 silencing, 426
 Homoeologous transcript levels, 169
Houzingenia, 9
 HS-CSO, *see* High stearic cottonseed oil
 Hsu, C.L., 190
 Hu, R., 467
 Huang, J., 467
 Hughes, E., 298

Hutchinson, J. B., 27, 33, 73
 Hutmacher, R. B., 473
 Huynh, T. T., 198
 Hybrid-B gene pool, 59

I

indicum, 73
 Induced mutation, genetic variability in
 crops by, 210
 Initiative 2000, 488
 Innes, N. L., 321
 Insertional mutagenesis
 En/I system, 218–219
 insertional mutagen
 TEs, *see* Transposable elements
 LTR-retrotransposons, 219
 In situ conservation, 64
 Inter-Simple Sequence Repeat
 inter- and intraspecific polymorphisms, 124
 microsatellite-complementary
 oligonucleotides as primers, 124
 Inter-specific hybridization, 87
 Iqbal, M. J., 39
 Islam-Faridi, M. N., 271–272
 ISSR, *see* Inter-Simple Sequence Repeat

J

Jenkins, J. N., 300
 JGI Web tool, 248
 Ji, Y. F., 271–272
 Jia, S. R., 466
 Jiang, B., 195
 Jiang C. X., 302
 Jin, W. J., 466
 Jumel, 56

K

Kanwar, V., 468
 Kareiva, P., 459
 Karnak, 57
 Katterman, F. R., 210
 Kearney, T. H., 57
 ketoacyl-ACP synthase III (KASIII), 362
 K-genome species
 seed dispersal in, 15
 taxonomy of, 8
 Khadi, B. M., 75
 Kinetin, 188
 KNAT transcription factors, 405
 Knight, R. L., 33
 Kohel, R. J., 10

KORRIGAN protein, 408
 Koshinsky, H. A., 198
 Kulkarni, V. N., 75
 Kumar, S., 196

L

Lacape, J. M., 27, 33, 105
 Latifolium cotton, 33
 Lavekar, R. C., 468
 Lennon, A. M., 197
 Li, W. D., 465
 Lin, H., 467
 LINE-like elements, 145
 Lint fibers, 440
 Liu, Q., 198
 Long fibers, 284
 Longley, A. E., 434
 Long limb type, 31
 Long oligonucleotide probes, 171–172
 LTR-retrotransposons, 219
 Lu, Y., 298
 Lubbers, E. L., 40
 Lyznik, L. A., 200

M

Maize TILLING Project, 215
 Manhardt, N. M., 463
 Marechal, R., 82
 Marker assisted selection (MAS), 101
 Martinell, B. J., 194
 Marvier, M., 459
 Mass selection, 86
 Mathio cotton, 74
 Mauer, F. M., 54
 May, L., 299
 May, O. L., 38–39, 40
 McCabe, D. E., 194
 McCallum, C. M., 214
 McCarty, J. C. Jr., 300
 McCreedy, C., 459
 McKnight, T. D., 270, 271–272
 Medberry, S. L., 200
Medicago truncatula, Tnt1 insertions
 in, 219
 Men, X., 464
 Menzel, M. Y., 263
 Mepiquat-based plant growth regulators
 (PGR), 42–43
 Meredith, W. R. Jr., 291, 298, 300
 Microarray development, *Gossypium*
 transcriptome analysis

cDNA microarray paper, 171
 cell elongation and primary cell wall
 formation, 176–178
 fiber initiation, 174–176
 fiber transcriptome, 173–174
 oligonucleotide microarray, 171–172
 secondary cell wall formation and
 maturation, 178
 short oligonucleotide probe, 172
 Microfibril, 385
 Mit Afifi cultivar, 61
 Mocó cotton, 33
 Modern improved germplasm pools, origins
 and development of
 American Pima cottons
 gene pool, sources of, 59
 Pima S-2, 59
 Pima S-3, 60
 productivity and fiber quality, 60
 progenitors of, 58, 59
 SxP, 58
 in Australia, 62
 Egyptian cottons
 acclimatizing, 58
 concern in early improvement efforts
 of, 57
 fiber quality of, 57–58
 Jumel and Ashmouni, 56
 selection and improvement, 57
 in Israel, 61
 Sea Island cottons
 fiber of, 54–55
 Fusarium resistant cultivars, 56
 improvement efforts in, 55–56
 Tanguis cottons, 61
 Molecular breeding activities, 330–331
 Molecular cytogenetics
 in cotton
 agronomically important genes,
 physical mapping of, 272–273
 chromosome structural abnormalities
 identification, 272
 chromosome translocations, 271
 rDNA, mapping of, 270–271
 total genomic DNA, 274–275
 on extended DNA fibers, 276
 principles of, 269–270
 Molecular markers
 based linkage maps in cotton
 interspecific population (*G. hirsutum* x
 G. barbadense), 103, 104–105
 intraspecific population (*G. hirsutum* x
 G. hirsutum), 104, 106

- and their advantages and disadvantages, 110–111
 - Molecular phylogenetic analyses, 5, 11
 - Mungari cotton, 75
 - Mutator-like sequences, 145
- N**
- N-Acylethanolamines (NAEs), 373
 - Na⁺/H⁺ antiporter, 346
 - Nanofibril, 385–386
 - Naqi, S., 263
 - Naranjo, 462
 - Narayanan, S. S., 89
 - Narula, A. M., 75
 - National Bureau of Plant Genetic Resources (NBPGR), 79
 - National Center for Biotechnology Information (NCBI), 227, 229
 - National Cotton Council, 358
 - NAU/RAPD/Rf13₁₄₈₀ and NAU/RAPD/Rf15₇₁₀, 113
 - NBS-LRR class of resistance genes, 329–330
 - Ndungo, V., 82
 - Nimblegen, 172
 - Noyer, J. L., 27
 - NSF project IOB-0090920, 329
- O**
- Old World, *G. arboreum* distribution in, 73–75
 - Oligo-based array, 171–172
 - Oliver, M. J., 200
 - Olson, P. D., 71
 - Ontologies, 228
 - used by *Gossypium* resources, 233
 - Opened bolls, shattering of locules from, 89
 - Organic evolution, 483
 - Osmotic adjustment, 339–340
 - Ow, D. W., 200
- P**
- Palmitic acid, altered levels of
 - in HO-CSO and HS-CSO, 368
 - KASII activity, 369
 - Paracel Transcript Assembler, 163
 - Patange, N. R., 468
 - Paterson, A. H., 40, 299, 423
 - Pawar, V. M., 468
 - PCR based marker system, 108, 115
 - Percival, A. E., 27
 - Peterson, D. E., 471
 - Phosphatidylcholine (PC), 362
 - Photoperiod neutral flowering, 55
 - Phylogenetic analysis of growth rates,
 - F-genome/A-genome lineage, 15
 - Phylogenetic history, of *Gossypium*, 12–13
 - Phytopathology, 314
 - Pilose hairiness trait, 113
 - Pima S-2, 59
 - Pima S-3, 60
 - Pima S-1 gene pool, 58–59
 - Plant breeding, 34
 - Plant Genome Mapping Laboratory (PGML), 151, 236
 - Plant germplasm variations, techniques applied to study, 102
 - Plant protoplast culture, 191
 - Plant responses to abiotic stress
 - osmotic adjustment, 339–340
 - transpiration, 340
 - water use efficiency, 339
 - Plant retrotransposons, transcription of, 219
 - Polyploid cottons
 - expression levels of homoeologous genes in, 424–425
 - genetic control of fiber growth and development in, 302
 - genome evolution from BAC sequencing, 421
 - origin and diversification of, 9–11
 - Polyploidy, 168
 - expression changes and silencing of genes duplicated by
 - abiotic stress conditions, homoeologous genes, 425–426
 - allopolyploidy, 426–427
 - organ-specific expression biases, homoeologous genes, 424–425
 - fates of genes duplicated by
 - rDNA genes, concerted evolution of, 422–423
 - sequence rate acceleration, in D-Genome, 423
 - genomic additivity and epigenetic stasis, 422
 - Gossypium* genome evolution, 421
 - repetitive elements, 421
 - POM1/AtCTL1* mutation, 405
 - Pray, C., 467
 - Price, H. J., 270, 271–272
 - Program for Annotating and Viewing ESTs (PAVE), 162–163
 - Putative QTLs

Putative QTLs (*cont.*)

- resistance to *Thielaviopsis basicola*, 323–324
- resistance to *Verticillium dahliae*
 - chromosome 5, 328–329
 - chromosome 11, 327–328, 329
- resistance to Xcm, 320
 - Xcm Races 2 and 4, 322
 - Xcm Races 7 and 18, 321

Q

Qin, M., 200

QTL mapping

- chromosome 6 for dense leaf pubescence, 108
- for fiber quality
 - fiber fineness/maturity, 298–300
 - fiber length, 296–298
 - fiber strength, 304
 - orthologous genomic regions, 303
 - phenotypic effects, 305
 - segregating populations, 293
- fiber quality traits, 112–113
- requirements to perform, 292
- RFLP genetic linkage map, cross of *G. hirsutum*, 107
- using SSRs
 - for fiber length, 124
 - for fiber strength, 123–124
- validation approaches, 303–304

R

Radiation mutagenesis, 210

Rajab, M., 27

Random amplified polymorphic DNA (RAPD)

- advantages of, 109
- applications
 - in CLCuD resistance, 114–115
 - in fertility restoration, 113
 - fiber quality traits analysis, 112–113
 - in phylogenetic and diversity studies, 111–112, 115
 - in physico-morphological traits studies, 113–114
- basic principle of, 109
- CLCuD epidemics, 111
- UBC111₃₀₀₀ and UBC188₅₀₀, 115
- RAPD-derived SCAR markers, 115
- Raska, W. A., 271–272
- Rathod, K. S., 468

rDNA genes

- concerted evolution of, 422–423
- mapping of, 270–271
- molecular cytogenetic studies of, 271
- repeated sequences, physical mapping of, 270–271

Refined cotton seed oil, 355

Regetz, J., 459

Reinisch, A. J., 106

Repeated sequences, physical mapping of rDNA, 270–271

- tandem arrays and dispersed repeats, 270

Repetitive DNA, 142–143

Repetitive elements

- movement due to polyploidization, 422
- in tetraploid cotton
 - A-genome specific, 146
 - categorization of, 144–145
 - tandem repeats, 145

Resistance gene analogues, 330

Resistance genes, classification of, 329

Restorer-specific markers, 113

Restriction fragment length polymorphism

- applications of
 - gene tagging and qtl mapping, 107–108
 - genetic diversity, *G. hirsutum* cultivars, 103
 - genetic linkage maps, interspecific population, 103, 106
 - genetic linkage maps, intraspecific population, 106–107
 - in cotton, 103
 - DNA fragments, 102
 - main steps involved in, 102
- Retrotransposons, 147
- LTR-retrotransposons, 219
- Reverse genetics
- definition of, 213
 - using TILLING
 - CEL I, 214–215
 - chemical mutagens, 214
 - for crop species, 215–216
 - dHPLC, mismatched DNA fragments identification by, 214
- RFLP, *see* Restriction fragment length polymorphism
- RGAs, *see* Resistance gene analogues
- R gene family, 118
- RHD3 and KOBITO, 409
- Ribosomal RNA genes, *see* rDNA genes
- Ring spinning, 285

- RNAi-ghFAD2-1 lines
 control levels of oleic and stearic acids through, 367–368
 high-oleic cotton seed oil derived from, 363–364
Rong, J., 301
Rotor spinning, 285
Rozelle, S., 467
Ruberson, J. R., 458, 460
- S**
Saha, S., 300
Sakel, 58
Saturation mutagenesis, 213
Saunders, J. H., 321
Schnabel, A., 5, 71, 422
Sea Island cottons
 fiber of, 54–55
 Fusarium resistant cultivars, 56
 improvement efforts in, 55–56
Secondary cell wall formation, functional genomics studies of, 178
Secondary wall deposition
 cotton fiber secondary wall
 transcriptome analysis for *Arabidopsis* homolog, 399–401
 endogenous controls for qPCR analysis, 397–398
 gene expression at 10 DPA, 400–401
 genes expression at 20 DPA vs. 6 DPA and 10 DPA, 396
 genes up-regulated at 24 DPA, 397
 GhCTL1 expression, 398–399
 function of genes related to
 carbon supply to cellulose synthetic enzymes, 407–408
 CESA isoforms, 404
 COBRA gene, 404–405
 cotton homolog of At2g37090, 406–407
 CSCs, 408
 GhCESA4 gene, 403
 G.h.fbr-sw sequence, 410
 homologous genes to *Arabidopsis* FLA11 and FLA12, 409
 KNAT transcription factors, 405
 KORRIGAN protein, 408
 POM1/*AtCTL1* mutation, 405
 RHD3 and KOBITO, 409
 SPIRAL1-LIKE genes, 406
 transcriptional control of, 405
Seed and trichome size, 13–14
Seed cotton yield, genetic gains for, 87–88
Seed coverings, 7
Seed dispersal, 15
Seeds, evolutionary history of, 13
 cellulose and suberin, 15
 fiber on seed coat, 14
 hairless, 14–15
Seelanan, T., 5, 71, 422
Semi-cluster type and cluster type, 30
Senchina, D. S., 13
Sesquiterpenes, biosynthesis of, 199
Sharma, O. P., 468
Shattuck, V., 210
Shen, X. L., 304
Shintani, D., 373
Short fibers, 284
Short oligonucleotide probes, 172
Silow, R. A., 73
Silver staining, 109
Simple-sequence repeat
 linkage mapping using, 122
 modification of, 124
 multiplexing of, 120
 phylogenetic and diversity studies using, 120
 EST-SSRs, 122
 polymorphism of species of *Gossypium*, 121–122
SINE-like elements, 145
Singh, A., 468
Singh, P., 89
Single-gene resistance, 318
Single nucleotide polymorphisms
 applications of
 in *Arabidopsis*, 125
 single-base changes and indels, 126
 genotyping techniques, 127
 as source of polymorphisms, 125
Single stranded conformational polymorphism (SSCP), 126–127
Sisterton, M. S., 463
Skovsted, A., 261–262, 434
Small, R. L., 423
Smith, R. H., 464
SNPs, *see* Single nucleotide polymorphisms
Spinning methods and fibe quality, 40
SPIRAL1-LIKE genes, 406
SSR arrays, variation in, 119
SSR polymorphism, 119
SSRs, *see* Simple-sequence repeat
Stelly, D. M., 270, 271–272, 300
Stephens, S. G., 28, 34, 37, 55
Stewart, J. M., 25–26, 46

- Stewart, J. Mc. D., 71, 190
 Stone, G. D., 473
 Sudweeks, E. M., 371
 Suppressive subtractive hybridization library, 178
- T**
- Tabashnik, B. E., 463, 474
 Tandemly repetitive DNA element families, 146
 Tandem repeats
 bidirectional homogenization of, 147
 evolution in tetraploid cotton, 146
 identification using Tandem Repeat Finder, 145
 Tanguis cottons, 61
 Tanwar, R. K., 468
 Targeting Induced Local Lesion IN Genomes, *see* TILLING
 Tetraploid cotton
 abiotic stress resistance of
 arid conditions, 343
 GhNHX1, 346
 heat and drought stress, 341, 342
 irrigation regimes, 343
 stomatal conductance and transpiration, 342
 classical cytogenetic studies of, 261–262
 reassociation kinetic analysis of, 143–144
 repetitive elements in
 A-genome specific, 146
 categorization of, 144–145
 tandem repeats, 145
 R-genes, disease resistance, 330
 Thermo-sensitive male sterility, 88
 TIGR Gene Index Cluster (TGICL), 164
 TILLING, 127, 375
 CEL I, 214–215
 chemical mutagens, 214
 for crop species, 215–216
 dHPLC, mismatched DNA fragments identification by, 214
 Ting, I. P., 190
 Tissue culture
 callus formation, gymnosperm and angiosperm pollen grains, 189
 of cotton
 callus formation, 189–190
 cell suspension cultures, 190
 organogenesis, 193
 somaclonal variation, 192
 somatic embryogenesis, 190–191
 transformation, 191, 192–193
 organogenesis, 189
 undifferentiated callus tissue, 188–189
 Tissue explants, 189
 Tnt1 system, 219
 Torres, J. B., 458, 460
 Totipotency, 188
 Townsend, B. J., 199
 Trans fatty acids (TFA), 356
 Transgenic cotton, 454
 advances of, 474–475
 ecological impacts of
 Bt cotton, 456
 herbicide resistance, 456–457
 potential for gene flow from, 472–473
 sociopolitical considerations of, 473–474
 See also Bt cotton
 Translocation breakpoints, 263
 Transposable elements
 amplification of, 147
 gene tagging by, 216
 in maize, 216–217
 Trichomes
 evolutionary history of, 15–16
 size and morphology in wild species, 13–14
 Trolinder, N. L., 190–200
 TropGENE-DB, 236
 TropGENE-DB (TGDB), 234–235
 Turcotte, E. L., 10
 Tyler, F. J., 29–30
- U**
- UBC111₃₀₀₀ and UBC188₅₀₀, 113, 115
 Ulian, E. C., 194
 Ulloa, M., 297, 300–302
 Undifferentiated callus tissue, 189
 United States Cotton belt groupings, 29
 acalas, 32
 boll weevil invasion and, 31–32
 early type cultivars, 31
 eastern big boll type, 30
 long limb type, 31
 semi-cluster type and cluster type, 30
 Unusual fatty acids, transgenic production of, 369
 from *C. palaestina*, 370
 DGAT activity and, 371
 Upland cotton, cytogenetic history of, 45

V

Van Esbroeck, G. A., 39

Van Haaren, M., 200

Vitamin E, cotton seed, 372–373

W

Wagad/Kalyan types, 75

Wang, G. L., 63

Ware, J. O., 30–31

Water deficiency, 338–339

Water stress and cotton productivity,
107–108

Water use efficiency, 339

Web resources, *Gossypium* data, 248

biochemical pathways, 252

crop science, 249–250

diversity and evolution, 250–251

maps, markers and genomic sequence, 249

proteome related, 250

transcriptome related, 251

See also Gossypium bioinformatics

resources

Wendel, J. F., 5, 10, 27–28, 37, 39, 71, 73,
422, 423, 426

Whitehouse, M. E. A., 463

Wild-type fiber cell, 15

Willems, K. R., 196

Wilson, L. J., 463

Wright, R. J., 321

Wu, K. M., 465

Wu, Y., 175

WUE, *see* Water use efficiency**X***Xanthomonas campestris* pv. *malvacearum*,
320

Xiang, X. T., 78–79

Xu, G., 465

Xue, D., 465

Y

Yang, C., 328

Yarn-spinning technologies, 285

Yseleta B2B6 resistant line, 62

Yu, J., 305

Yu, L., 458

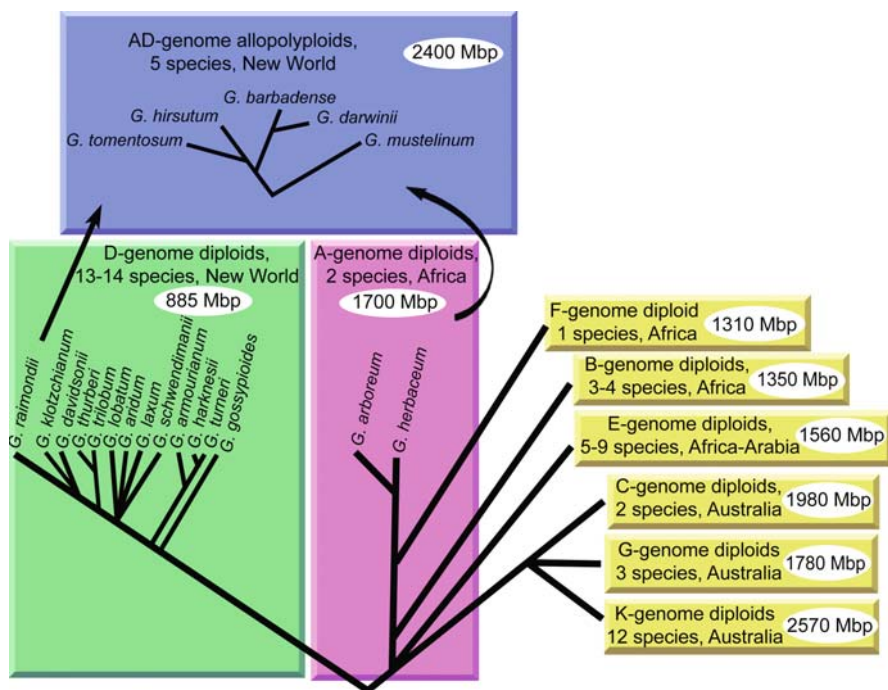
Z

Zhang, J. F., 298

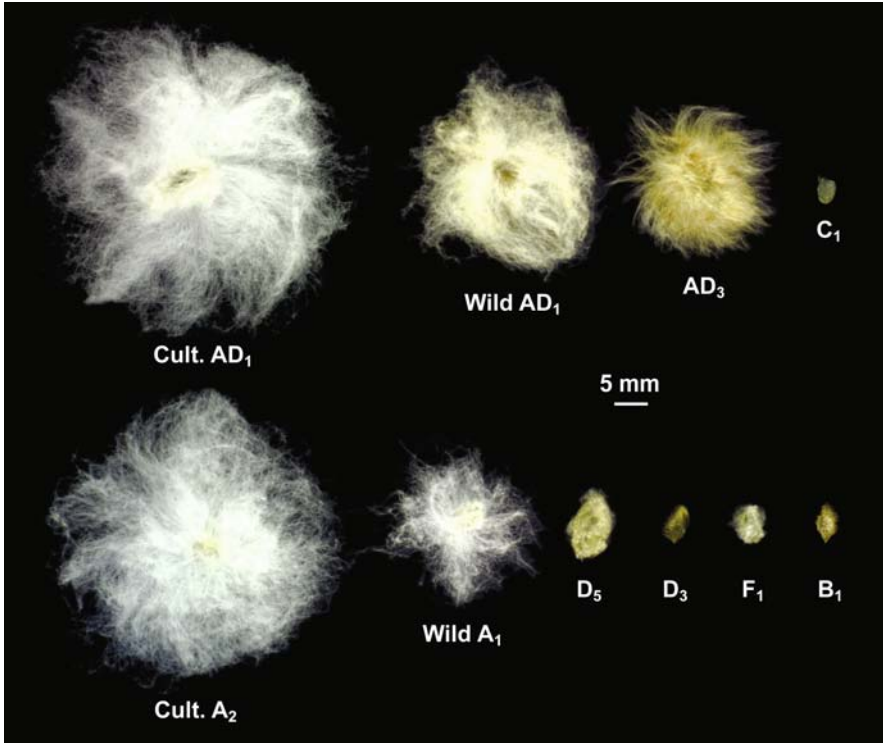
Zhao, X., 423

Zhao, X. -P., 421

Zhou, G. -Y., 195



Chapter 1, Fig. 1 Evolutionary history of *Gossypium*, as inferred from multiple molecular phylogenetic data sets. The closest relative of *Gossypium* is a lineage containing the African-Madagascan genus *Gossypoides* and the Hawaiian endemic genus *Kokia*. Following its likely origin 5–10 mya, *Gossypium* split into three major diploid lineages: the New World clade (D-genome); the African-Asian clade (A-, B-, E- and F-genomes); and the Australian clade (C-, G-, and K-genomes). This global radiation involved several trans-oceanic dispersal events and was accompanied by morphological, ecological, and chromosomal differentiation (2C genome sizes shown in white ellipses). Interspecific hybridization is implicated in the evolution of approximately one-fourth of the genus. Allopolyploid cottons formed following trans-oceanic dispersal of an A-genome diploid to the Americas, where the new immigrant underwent hybridization, as female, with a native D-genome diploid similar to modern *G. raimondii*. Polyploid cotton probably originated during the Pleistocene (1–2 mya), with the five modern species representing the descendants of an early and rapid colonization of the New World tropics and subtropics



Chapter 1, Fig. 2 Representative seed and trichome diversity in *Gossypium*. Seed and trichome size and morphology are exceedingly variable in the genus. Most wild species have relatively small seeds (<5 mm in any dimension) with equally short fibers. Long (spinnable) fiber evolved only once, in the ancestor of modern A-genome cottons, which subsequently donated this capacity to modern tetraploid species, including the commercially important *G. hirsutum* and *G. barbadense*, at the time of allopolyploid formation in the mid-Pleistocene. See text for additional detail. Key to species: Cult. AD₁ = *G. hirsutum* TM1; Wild AD₁ = *G. hirsutum* Tx2094 from the Yucatan Peninsula; AD₃ = *G. tomentosum* WT936 from Hawaii; C₁ = *G. sturtianum* C₁-4 from Australia; Cult. A₂ = *G. arboreum* AKA8401; Wild A₁ = *G. herbaceum* subsp. *africanum* from Botswana; D₅ = *G. raimondii* from Peru; D₃ = *G. davidsonii* D_{3d}-32 from Baja California; F₁ = *G. longicalyx* F₁-3 from Tanzania; B₁ = *G. anomalum* B₁-1 from Africa



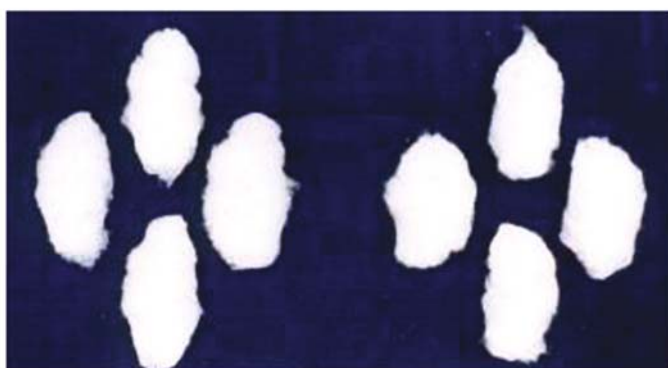
Chapter 4, Fig. 4 Dhummad variety of *G. herbaceum* having closed bolls cultivated in coastal salinity of Gujarat, India



G. arboreum

G. hirsutum (Abadhita)

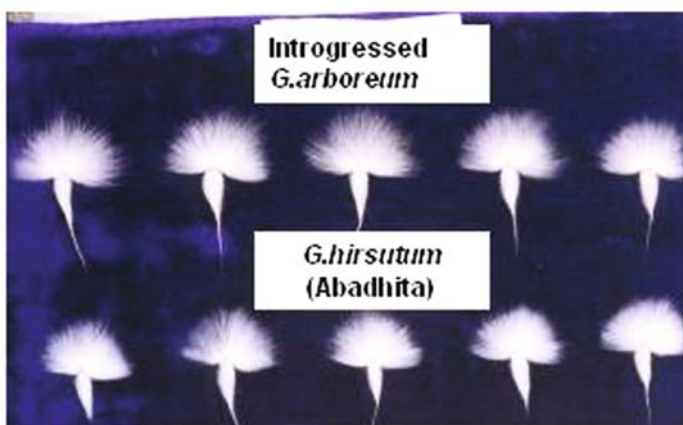
Opened bolls



G. arboreum

G. hirsutum (Abadhita)

Locules



Introgressed
G. arboreum

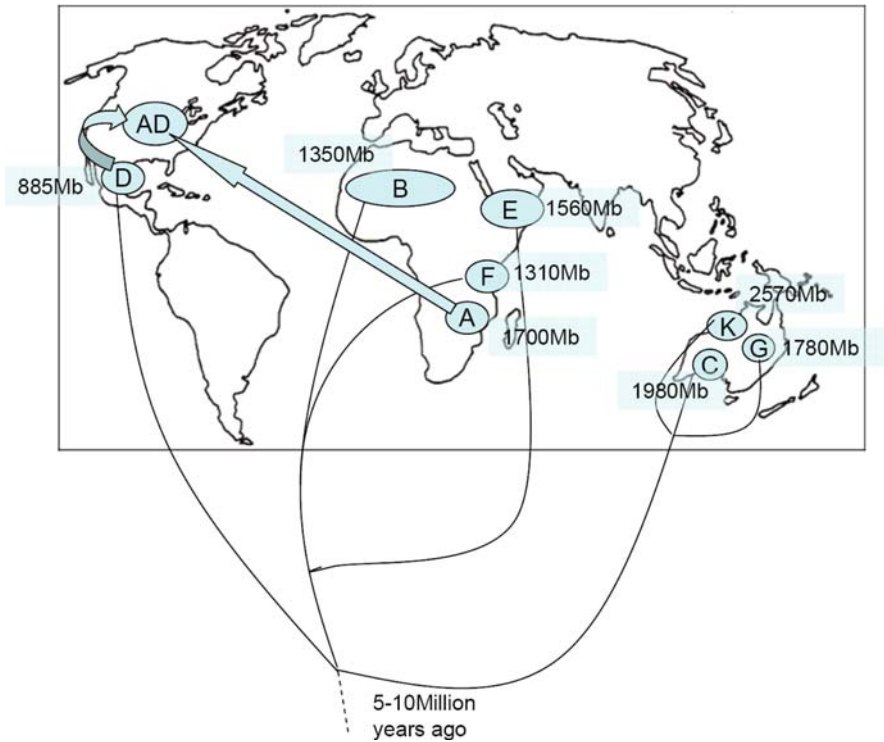
G. hirsutum
(Abadhita)

Halo length

Chapter 4, Fig. 5 Comparison of introgressed *G. arboreum* (backcross derivatives of cross between *G. arboreum* and *G. hirsutum*) with *G. hirsutum* for (a) boll size (b) locule size and (c) halo length

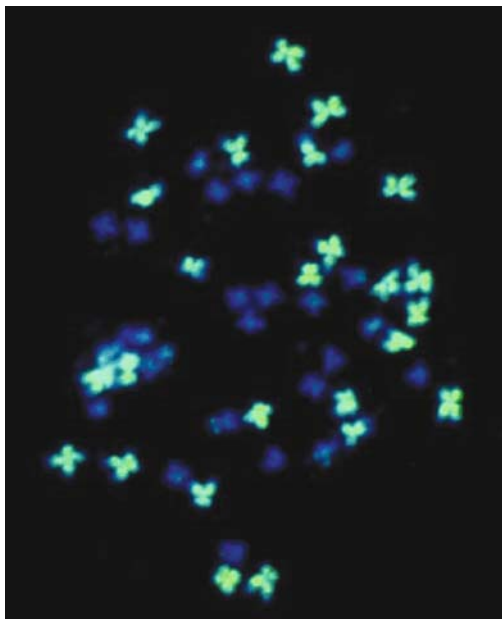


Chapter 4, Fig. 6 Improved seed size in introgressed *G. arboreum* lines (AH 56 and AH 29) in comparison with *G. arboreum* parent (PA 140), *G. hirsutum* parent (Poornima) and small seeded *G. arboreum* line

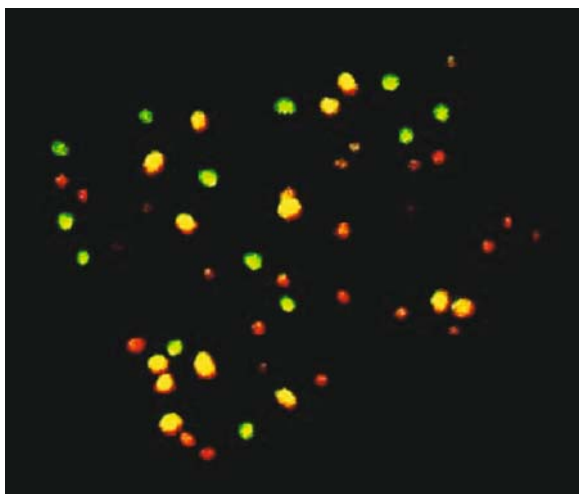


Chapter 6, Fig. 1 The genome size and evolutionary relationship among different cotton species. Modified from <http://www.eeob.iastate.edu/faculty/WendelJ/images/map2.jpg>

Chapter 11, Fig. 3 Use of GISH to differentiate the A subgenome (in green) and the D subgenome (in blue) of *G. hirsutum*. Total DNA of *G. herbaceum* (A1 genome) labeled with digoxenin was used as probe; and total DNA of *G. thurberi* was used as blocking DNA



Chapter 11, Fig. 4 Use of GISH to detect the presence of *G. longicalyx* chromosomes in mitotic plates of the trispecific hybrid $[(G. hirsutum \times G. thurberi)^2 \times G. longicalyx]$. Total DNA of *G. longicalyx* (labeled with digoxenin) and *G. hirsutum* (labeled with biotine) were used as probes. *G. longicalyx* genetic material is revealed in green



Chapter 11, Fig. 5 Detection of recombinations (arrows) in a BC2 plant of the trispecies hybrid $[(G. \textit{hirsutum} \times G. \textit{thurberi})^2 \times G. \textit{longicalyx}]$. Total DNA of *G. longicalyx* (labeled with digoxenin) and *G. hirsutum* (labeled with biotine) were used as probes. *G. longicalyx* genetic material is revealed in green

